Materialien zum Modul *Formale Methoden der Sozialwissenschaften*
Teil I: Begriffsbildungen und Modelle

G. Rohwer

Version 1 **August 2007**

Fakultät für Sozialwissenschaft
Ruhr-Universität Bochum, GB 1
44780 Bochum

goetz.rohwer@ruhr-uni-bochum.de

tm.tex August 2007

Dieser Text enthält Materialien zum Modul *Formale Methoden der Sozialwissenschaft*.

Hinweise zum Text

- Wie im Inhaltsverzeichnis angegeben wird, gliedert sich der Text in Kapitel, die meisten von ihnen auch in Abschnitte. Eine weitere Untergliederung in Paragraphen wird zu Beginn jedes Kapitels angegeben.
- Einfache Anführungszeichen werden zur Kennzeichnung sprachlicher Ausdrücke verwendet, doppelte Anführungszeichen werden verwendet, um Zitate kenntlich zu machen oder um anzudeuten, dass ein Ausdruck unklar ist und/oder metaphorisch verwendet wird. Innerhalb von Zitaten wird versucht, die im Original verwendeten Anführungszeichen zu reproduzieren. Wenn nicht anders angegeben, folgen Hervorhebungen in Zitaten stets dem Original; eigene Zusätze, Änderungen und Auslassungen werden durch eckige Klammern kenntlich gemacht.
- Wir unterscheiden die Zeichen '=' und ':='. Ein Gleichheitszeichen mit vorangestelltem Doppelpunkt wird verwendet, um anzudeuten, dass eine definitorische Gleichsetzung vorgenommen wird, d.h. der Ausdruck auf der linken Seite wird durch den Ausdruck auf der rechten Seite definiert. Dagegen dient ein einfaches Gleichheitszeichen zur Formulierung einer Gleichheitsbehauptung und setzt deshalb voraus, dass beide Seiten schon definiert sind.
- Als Dezimalpunkt wird ein Punkt und nicht, wie im Deutschen üblich, ein Komma verwendet.
- Bei den Notationen aus der Mengenlehre und zum Funktionsbegriff folgen wir den Ausführungen bei Rohwer und Pötter (2001, S. 21ff.).

Inhalt

Kapitel 1

Statistische Begriffsbildungen

In diesem Kapitel wird mit der Diskussion statistischer Begriffsbildungen begonnen. Zunächst geht es im ersten Abschnitt darum, den theoretischen Ansatz statistischer Begriffsbildungen zu verstehen; dann folgen Überlegungen zu statistischen (Sozial-) Strukturbegriffen; schließlich werden einige Grundbegriffe der statistischen Regressionsrechnung besprochen.

1.1 Statistische Variablen

1. Bezugnahme auf Gesamtheiten. Die Entwicklung der Statistik kann als eine Folge des Wunsches verstanden werden, empirisch explizierbare Vorstellungen über Gesamtheiten zu gewinnen, die nicht unmittelbar überschaubar sind. Ursprünglich ging es in erster Linie um eine Erfassung von Bevölkerungen (Populationen). Etwa seit der zweiten Hälfte des 19. Jahrhunderts hat sich die Statistik zu einer abstrakten Methodenwissenschaft entwickelt, deren Begriffsbildungen auf beliebige Gesamtheiten anwendbar sind. Dass in irgendeiner Weise eine Bezugnahme auf Gesamtheiten erfolgt, ist jedoch in jedem Fall relevant, um statistische Aussagen zu verstehen. Maurice Kendall und Alan Stuart haben das zu Beginn ihrer „Advanced Theory of Statistics“ (1977: 1) so ausgedrückt:

„The fundamental notion in statistical theory is that of the group or aggregate, a concept for which statisticians use a special word – “population”. This term will be generally employed to denote any collection of objects under consideration, whether animate or inanimate; for example, we shall consider populations of men, of plants, of mistakes in reading a scale, of barometric heights on different days, and even populations of ideas, such as that of the possible ways in which a hand of cards might be dealt. [...] The science of Statistics deals with the properties of populations. In considering a population of men we are not interested, statistically speaking, in whether some particular individual has brown eyes or is a forger, but rather in how many of the individuals have brown eyes or are forgers, and whether the possession of brown eyes goes with a propensity to forgery in the population. We are, so to speak, concerned with the properties of the population itself. Such a standpoint can occur in physics as well as in demographic sciences.“

In diesem Zitat wird auch schon darauf hingewiesen, dass sich statistische Aussagen *in spezifischer Weise* auf Gesamtheiten beziehen; das wird weiter unten (in § 6) genauer besprochen. Bereits an dieser Stelle kann aber festgestellt werden, dass Aussagen über Gesamtheiten von Aussagen über ihre individuellen Mitglieder zu unterscheiden sind.

2. Gesamtheiten als Mengen. Wenn in der Statistik von Gesamtheiten gesprochen wird, sind Mengen im Sinne der Mengenlehre gemeint, d.h. Zusammenfassungen von Elementen zu einer gedanklichen Einheit, wobei von allen möglicherweise vorhandenen Beziehungen zwischen den Elementen abstrahiert wird. Der Begründer der Mengenlehre, Georg Cantor, hat einmal folgende Definition gegeben:

„Unter einer „Menge“ verstehen wir jede Zusammenfassung M von bestimmten wohlunterschiedenen Objekten unsrer Anschauung oder unseres Denkens (welche die „Elemente“ von M genannt werden) zu einem Ganzen.“ (Cantor 1962: 282)

Der gedankliche Ansatz ist allgemein und abstrakt. Es gibt keinerlei Einschränkungen hinsichtlich der Arten von Objekten, die man gedanklich zu einer Menge zusammenfassen kann. Es muss sich auch nicht unbedingt um materielle Objekte im umgangssprachlichen Sinn dieses Worts handeln, zum Beispiel können auch Zahlen, Eigenschaften, Ereignisse und Gebiete eines Raums zu Mengen zusammengefasst werden, und auch Mengen selbst können wiederum als Elemente zur Definition neuer Mengen verwendet werden.

Wichtig ist auch, dass mit dem Mengenbegriff nur eine gedankliche Einheit der jeweils in Betracht gezogenen Elemente gemeint ist (von der Frage, ob und ggf. in welcher Weise der Menge auch eine „reale Einheit“

entspricht, wird also abgesehen). Ebenfalls wird von allen Beziehungen abstrahiert, die möglicherweise zwischen den zu einer Menge zusammengefassten Elementen bestehen bzw. hergestellt werden können. Infolgedessen wird auch von räumlichen oder zeitlichen Anordnungen der Elemente, soweit man ggf. davon sprechen kann, abgesehen. Insbesondere spielt die Reihenfolge, in der die Elemente einer Menge vorgestellt oder aufgeschrieben werden, keine Rolle.[1]

3. Repräsentation von Gesamtheiten. Im Rahmen statistischer Überlegungen werden Gesamtheiten immer als Mengen aufgefasst, und wenn in diesem Text ohne weiteren Zusatz von Gesamtheiten gesprochen wird, sind deshalb stets Mengen gemeint. Bei ihren Elementen kann es sich um real existierende oder um fiktive Objekte handeln. Im ersten Fall gibt es die Objekte in der menschlichen Erfahrungswelt (einschließlich der empirisch zugänglichen Vergangenheit), im zweiten Fall gibt es sie nur in der Vorstellungswelt eines oder mehrerer Menschen. Dementsprechend kann man im ersten Fall von *empirischen*, im zweiten Fall von *fiktiven Gesamtheiten* sprechen. Fiktive Gesamtheiten können endlich oder unendlich viele Elemente enthalten, empirische Gesamtheiten haben jedoch immer nur endlich viele Elemente. Weiterhin gilt natürlich auch, dass empirische Gesamtheiten nur Elemente enthalten können, die es in der *bisherigen* Erfahrungswelt von Menschen tatsächlich gibt oder gegeben hat, also insbesondere keine möglicherweise in der Zukunft existierenden Objekte.

In der *Sozialstatistik*[2] beschäftigt man sich mit empirischen Gesamtheiten. Dabei ist zu berücksichtigen, dass man sich oft nicht unmittelbar und vollständig auf alle Elemente einer intendierten empirischen Gesamtheit

[1] Dies muss auch deshalb betont werden, weil in der soziologischen Literatur (und in der Umgangssprache) der Mengenbegriff gelegentlich anders verwendet wird. Zum Beispiel schreiben R. Boudon und F. Bourricaud in ihren „Soziologischen Stichworten“ (1992: 184): „Eine Menge und eine Masse sind nicht dasselbe. Die beiden Bezeichnungen beziehen sich auf unterschiedliche soziale Situationen. In einer Menge, die einem Fußballspiel beiwohnt, stehen die Beteiligten in Interaktionsbeziehungen zueinander. Die einen pfeifen, die anderen klatschen; und in beiden Lagern entwickelt sich eine Solidarität sowie – je nach dem Grad ihrer Begeisterung – eine Differenzierung zwischen den Fans. [...] Die *Masse* derjenigen dagegen, die ein Fernsehprogramm verfolgen oder eine Zeitung lesen, hat kaum Möglichkeiten, miteinander in Kontakt zu treten. Außerdem kommen die Beziehungen zwischen ihnen nur durch die Vermittlung der ausgestrahlten Sendung oder der Druckseite zustande. Ihre Gemeinsamkeiten beschränken sich darauf, daß sie Leser derselben Zeitung oder Zuschauer desselben Programms sind.“ Offenbar wird hier von einer „Menge“ im Unterschied zu einer „Masse“ dann gesprochen, wenn es zwischen ihren Elemente gewisse Interaktionsbeziehungen gibt. Es ist natürlich zulässig, das Wort ‘Menge’ auch in dieser Bedeutung zu verwenden; es ist aber wichtig zu wissen, dass das Wort in der Mengenlehre und der sich an sie anschließenden Statistik anders verwendet wird, nämlich in einer Bedeutung, die von allen möglicherweise vorhandenen Beziehungen zwischen den Elementen abstrahiert.

[2] In der Literatur findet man gelegentlich die Wortkombination *Wirtschafts- und Sozialstatistik*. In diesem Text wird von Sozialstatistik in einem umfassenden Sinn gesprochen, der Bezugnahmen auf wirtschaftliche Sachverhalte einschließt.

beziehen kann. Als Beispiel kann man an die Gesamtheit der Menschen denken, die in Deutschland im September 2003 arbeitslos gewesen sind. Offenbar ist eine empirische Gesamtheit gemeint, die in diesem Beispiel aus Menschen besteht, die im angegebenen Zeitraum tatsächlich gelebt haben. Aber es ist auch klar, dass diese Gesamtheit nicht unmittelbar beobachtet werden kann. Deshalb ist man gezwungen, sich in irgendeiner Form eine Repräsentation der Gesamtheit, über die man sprechen möchte, zu verschaffen.

Mit Repräsentationen sind in diesem Zusammenhang gegenständliche oder sprachliche Hilfsmittel gemeint, die es erlauben sollen, sich die Elemente einer nicht unmittelbar überschaubaren Gesamtheit zu vergegenwärtigen. Je nach dem verfügbaren Vorwissen gibt es dafür unterschiedliche Möglichkeiten. Als Beispiel kann man an eine Kartei im Personalbüro eines Unternehmens denken, die für jede in dem Unternehmen beschäftigte Person eine Karteikarte mit Informationen über die Person enthält. Eine solche Kartei repräsentiert dann (im hier gemeinten Sinn) die Belegschaft des Unternehmens, wobei es gleichgültig ist, in welchen technischen Formen die Kartei existiert (etwa in Form eines Karteikastens mit Karteikarten oder in Gestalt einer Datei in einem Computer).

Eine minimale Anforderung an eine Repräsentation besteht darin, dass es für die Elemente der intendierten Gesamtheit Namen gibt, die ihre Unterscheidung ermöglichen. Solche Namen benötigt man auch dann, wenn eine empirische Identifikation noch gar nicht stattgefunden hat, sondern zunächst nur als Möglichkeit vorstellbar ist (wie zum Beispiel bei den Menschen, die im September 2003 in Deutschland arbeitslos gewesen sind). Folgende allgemeine Notation eignet sich sowohl für empirische als auch für fiktive Gesamtheiten: $\Omega := \{\omega_1, \ldots, \omega_n\}$.[3] In dieser Notation sind die Symbole $\omega_1, \ldots, \omega_n$ Namen der Elemente (wobei die natürliche Zahl n auf die Anzahl der Namen verweist), und die Mengenklammern geben an, dass sie (die Namen bzw. die durch sie repräsentierten Elemente) zu einer Menge zusammengefasst werden sollen. Schließlich erhält diese Menge per Definition den Namen Ω.

4. Statistische Variablen. Statistische Aussagen über Gesamtheiten gehen von deren Elementen aus. Die einfachste Aussage stellt nur fest, wieviele Elemente die Gesamtheit enthält. Alle weiteren statistischen Aussagen über Gesamtheiten gehen von Eigenschaften aus, die sich zunächst ihren Elementen zurechnen lassen. Solche Eigenschaften werden durch *statistische Variablen* repräsentiert, die allgemein als *Funktionen im mathematischen Sinn* definiert sind. Zur Erläuterung kann folgendes Schema verwendet werden: $X : \Omega \longrightarrow \tilde{\mathcal{X}}$. Hierbei ist Ω eine statistische Gesamtheit (wir sprechen oft von einer *Objektmenge* oder auch von einer *Referenzmenge*), und $\tilde{\mathcal{X}}$ ist der *Merkmalsraum* der Variablen, d.h. eine Menge von Attributen, so dass jedes Element von Ω durch genau eines dieser Attribute charakterisiert werden kann. Schließlich ist X der Name der statistischen Variablen, also der Funktion, die jedem Element von Ω das ihm entsprechende Attribut zuordnet.

[3]In diesem Text werden die Zeichen '=' und ':=' unterschieden. Ein Gleichheitszeichen mit vorangestelltem Doppelpunkt wird verwendet, um anzudeuten, dass eine definitorische Gleichsetzung vorgenommen wird, d.h. der Ausdruck auf der linken Seite wird durch den Ausdruck auf der rechten Seite definiert. Dagegen dient ein einfaches Gleichheitszeichen zur Formulierung von Gleichheitsbehauptungen und setzt deshalb voraus, dass beide Seiten schon definiert sind.

Als Beispiel kann man an eine statistische Variable denken, die jedem Mitglied einer Gesamtheit von Menschen sein Geschlecht zuordnet, also entweder das Attribut 'männlich' oder das Attribut 'weiblich'. Offenbar kann man diese Attribute auch durch Zahlen repräsentieren, also etwa einen Merkmalsraum $\tilde{\mathcal{X}} := \{0, 1\}$ verwenden und vereinbaren, dass die Zahl 0 das Attribut 'männlich' und die Zahl 1 das Attribut 'weiblich' bedeuten soll. In diesem Beispiel handelt es sich um einen *qualitativen* Merkmalsraum, womit gemeint ist, dass es für die Elemente des Merkmalsraums keine sinnvolle lineare Ordnung gibt. Dagegen sind *quantitative* Merkmalsräume dadurch definiert, dass es für ihre Elemente eine sinnvolle lineare Ordnung gibt; als Beispiel kann man an einen Merkmalsraum $\tilde{\mathcal{Y}} := \{0, 1, 2, 3, \ldots\}$ denken, dessen Elemente zur Feststellung des Alters von Menschen (in diesem Beispiel in vollendeten Lebensjahren) verwendet werden können.[4]

Es sei betont, dass statistische Variablen Funktionen sind und von *logischen Variablen* (Leerstellen in Aussageformen) unterschieden werden müssen.[5] Außerdem dürfen statistische Variablen nicht mit ihren Merkmalsräumen verwechselt werden, wie dies gelegentlich in der Methodenliteratur geschieht.[6] Man kann natürlich abkürzend ohne Zusatz von Variablen sprechen, wenn aus dem Kontext hervorgeht, ob statistische oder logische Variablen gemeint sind.

Zum Verständnis ist auch zu beachten, dass das Wort 'Funktion' in unterschiedlichen Bedeutungen verwendet werden kann. Hauptsächlich sind zwei Verwendungsmöglichkeiten zu unterscheiden. Einerseits eine Verwendung, in der das Wort 'Funktion' auf einen Zweck, eine Leistung oder eine Aufgabe verweisen soll; andererseits die mathematische Verwendung des Funktionsbegriffs, in der das Wort die Zuordnung der Elemente einer Menge zu Elementen derselben oder einer anderen Menge meint. In Teilen der soziologischen Literatur (insbesondere im Umkreis sogenannter „funktionalistischer“ Theorieansätze) wird das Wort in der ersten dieser beiden

[4]Ausführlichere Überlegungen zu unterschiedlichen Arten von Merkmalsräumen findet man bei Rohwer und Pötter (2002a, Kap. 4).

[5]Dazu ausführlich Rohwer und Pötter (2002b, Kap. 9).

[6]Unklare Verwendungen des Variablenbegriffs in der Methodenliteratur werden bei Rohwer und Pötter (2002a: 14ff.) besprochen.

Tabelle 1.1-1 Fiktive Daten für eine eindimensionale statistische Variable X (links) und eine zweidimensionale statistische Variable (X, Y) (rechts).

ω	$X(\omega)$
ω_1	0
ω_2	1
ω_3	0
ω_4	0
ω_5	1
ω_6	0
ω_7	1
ω_8	1
ω_9	0
ω_{10}	0

ω	$X(\omega)$	$Y(\omega)$
ω_1	0	22
ω_2	1	29
ω_3	0	26
ω_4	0	25
ω_5	1	26
ω_6	0	24
ω_7	1	22
ω_8	1	25
ω_9	0	25
ω_{10}	0	23

Bedeutungen verwendet;[7] wir werden das Wort in diesem Text jedoch ausschließlich in seiner mathematischen Bedeutung verwenden.[8]

Es sei auch angemerkt, dass eine Funktion im mathematischen Sinn nicht mit der Vorstellung eines „funktionalen Zusammenhangs" verwechselt werden darf. Bereits zur Interpretation statistischer Variablen passt eine solche Vorstellung offenbar nicht.

5. Mehrdimensionale statistische Variablen. In vielen Fällen ist es möglich und oft von besonderem Interesse, die Elemente einer Gesamtheit gleichzeitig durch zwei oder mehr Arten von Merkmalen zu charakterisieren. Man spricht dann von *mehrdimensionalen statistischen Variablen*, wobei jeder einzelne Merkmalsraum als eine „Dimension" (in einem rein formalen, nicht räumlich aufzufassenden Sinn) zählt. Denkt man zur Illustration wieder an eine Personengesamtheit, könnte jeder Person gleichzeitig ein Geschlecht und ein Alter zugeordnet werden. Dem entspricht dann eine zweidimensionale statistische Variable $(X, Y) : \Omega \longrightarrow \tilde{\mathcal{X}} \times \tilde{\mathcal{Y}}$, wobei sich der Merkmalsraum $\tilde{\mathcal{X}} := \{0, 1\}$ auf das Geschlecht und der Merkmalsraum $\tilde{\mathcal{Y}} := \{0, 1, 2, 3, \ldots\}$ auf das Alter bezieht. Dementsprechend wäre $(X, Y)(\omega) = (1, 25)$ so zu verstehen, dass ω der Name einer 25jährigen Frau ist.

[7] Zum Beispiel schreibt H. Joas in einer Einführung für ein Lehrbuch der Soziologie (2001: 21): „Der Ausdruck „Funktion" bezeichnet den Beitrag, den jede soziale Beziehung, Position, Organisation, jeder Wert oder jede Eigenschaft einer Gesellschaft für das soziale System als Ganzes leistet. [...] So besteht die Funktion von Schulen darin, Schüler auszubilden, die über die von den Unternehmen geforderten Fertigkeiten verfügen und am öffentlichen Leben als Bürger ihres Landes teilnehmen können." Eine Besprechung unterschiedlicher Verwendungsweisen des Funktionsbegriffs in der soziologischen Literatur findet man bei Merton (1957: 20ff.).

[8] In der Notation und Terminologie folgen wir den Ausführungen bei Rohwer und Pötter (2001: 24ff.). Zur Geschichte des mathematischen Funktionsbegriff vgl. man Steiner (1969).

Tabelle 1.1-1 illustriert die Begriffsbildungen mit fiktiven Daten. Die linke Hälfte illustriert eine eindimensionale, die rechte Hälfte eine zweidimensionale statistische Variable. Die Personengesamtheit ist in beiden Fällen identisch und besteht aus 10 Personen. Die eindimensionale Variable ordnet jeder Person ein Geschlecht zu, die zweidimensionale Variable ordnet jeder Person außerdem ein Alter zu.

6. Der statistische Verteilungsbegriff. Mit dem Begriff einer statistischen Variablen steht ein sehr allgemeines Schema zur Repräsentation von Objekten und ihrer Merkmale zur Verfügung. Kennt man eine statistische Variable $X : \Omega \longrightarrow \tilde{\mathcal{X}}$, kennt man auch für jedes Element $\omega \in \Omega$ den Merkmalswert $X(\omega)$. Das statistische Erkenntnisinteresse zielt jedoch gar nicht auf ein solches Wissen über die individuellen Mitglieder der jeweiligen Gesamtheit, sondern nur auf das Ausmaß, in dem bestimmte Merkmalswerte in der Gesamtheit vorkommen. Als Beispiel können die Daten für die Variable X in Tabelle 1.1-1 dienen. Aus statistischer Sicht interessiert nicht, dass ω_1 der Name einer männlichen und ω_2 der Name einer weiblichen Person ist, sondern dass es in der Gesamtheit sechs männliche und vier weibliche Personen gibt; oder in relativen Häufigkeiten ausgedrückt: 60 % sind männlich und 40 % sind weiblich.

Diesem spezifischen Erkenntnisinteresse dient der Begriff einer *statistischen Verteilung*.[9] Wie bei statistischen Variablen handelt es sich um Funktionen; aber – und darin kommt der statistische Perspektivenwechsel zum Ausdruck – als Definitionsbereich der Funktion dient jetzt nicht die Objektmenge Ω, sondern die Gesamtheit aller möglichen Merkmalskombinationen, also die Potenzmenge $\mathcal{P}(\tilde{\mathcal{X}})$ des Merkmalsraums $\tilde{\mathcal{X}}$.[10] Also kann folgende Definition gegeben werden: Die *Verteilung einer statistischen Variablen* $X : \Omega \longrightarrow \tilde{\mathcal{X}}$ ist eine Funktion, die jeder Teilmenge $\tilde{X}$ des Merkmalsraums $\tilde{\mathcal{X}}$ die (absolute oder relative) Häufigkeit derjenigen Objekte in Ω zuordnet, die einen Merkmalswert in $\tilde{X}$ aufweisen.

Zur Notation verwenden wir $\mathrm{P}^*[X]$, wenn auf absolute Häufigkeiten Bezug genommen wird, und $\mathrm{P}[X]$, wenn auf relative Häufigkeiten Bezug genommen wird. In eckigen Klammern steht der Name der Variablen, deren Verteilung bezeichnet werden soll.[11] Somit gelangt man zu den Definitionen $\mathrm{P}^*[X](\tilde{X}) := |\{\omega \in \Omega \,|\, X(\omega) \in \tilde{X}\}|$ und $\mathrm{P}[X](\tilde{X}) := \mathrm{P}^*[X](\tilde{X})/|\Omega|$, wobei $\tilde{X}$ eine beliebige Teilmenge von $\tilde{\mathcal{X}}$ ist.[12] Als Konvention wird ver-

[9] In gleicher Bedeutung spricht man auch von *Häufigkeits-* und *Merkmalsverteilungen* oder auch kurz von *der Verteilung* (einer statistischen Variablen).

[10] Wenn M irgendeine Menge ist, bezeichnet $\mathcal{P}(M)$ ihre *Potenzmenge*, d.h. die Menge aller Teilmengen von M.

[11] Diese eckigen Klammern bilden einen Teil des Namens der Funktion und dürfen nicht mit Argumenten verwechselt werden, die in runden Klammern angehängt werden. Natürlich kann die Angabe in den eckigen Klammern entfallen, wenn aus dem Kontext deutlich wird, auf welche Variablen Bezug genommen wird.

[12] Wenn M eine endliche Menge ist, soll $|M|$ die Anzahl ihrer Elemente bedeuten.

einbart, dass, wenn ohne Zusatz von Häufigkeiten gesprochen wird, stets relative Häufigkeiten gemeint sind. Dies soll analog auch für das Reden von statistischen Verteilungen gelten.

Zur Illustration beziehen wir uns wieder auf die Variable X in Tabelle 1.1-1. In diesem Beispiel ist der Merkmalsraum $\tilde{\mathcal{X}} = \{0,1\}$, es gibt also vier Teilmengen mit folgenden absoluten bzw. relativen Häufigkeiten:

$\tilde{X}$	$\mathrm{P}^*[X](\tilde{X})$	$\mathrm{P}[X](\tilde{X})$
$\emptyset$	0	0.0
$\{0\}$	6	0.6
$\{1\}$	4	0.4
$\tilde{\mathcal{X}}$	10	1.0

Es sollte beachtet werden, dass als Argumente einer Häufigkeitsfunktion $\mathrm{P}[X]$ nicht Elemente, sondern Teilmengen des Merkmalsraums der Variablen X verwendet werden; solche Teilmengen werden auch *Merkmalsmengen* genannt. Die Berücksichtigung der leeren Menge $\emptyset$ und der Gesamtmenge $\tilde{\mathcal{X}}$ dient natürlich nur der formalen Vollständigkeit.[13]

7. Statistische Aussagen über Gesamtheiten. Bereits zu Beginn dieses Abschnitts wurde betont, dass sich statistische Aussagen stets auf Gesamtheiten beziehen; jetzt kann genauer gesagt werden, dass es sich stets um *Aussagen über statistische Verteilungen* handelt. Dass es sich um eine spezifische Art von Aussagen über Gesamtheiten handelt, wird deutlich, wenn man darauf achtet, dass unsere Sprache zweideutig ist, wenn im Plural über die Mitglieder irgendeiner Gesamtheit gesprochen wird. Eine Aussage der

[13]Die Idee einer Häufigkeitsfunktion kann leicht für mehrdimensionale Variablen verallgemeinert werden. Als Beispiel verwenden wir die in Tabelle 1.1-1 angegebene Variable (X,Y), bei der sich X auf das Geschlecht und Y auf das Alter der Mitglieder einer aus 10 Personen bestehenden Gesamtheit Ω bezieht. Als Merkmalsmengen kommen jetzt alle Teilmengen des Merkmalsraums von (X,Y), also des kombinierten Merkmalsraums $\tilde{\mathcal{X}} \times \tilde{\mathcal{Y}}$, in Betracht. Die Häufigkeitsfunktion von (X,Y), für die die Notation $\mathrm{P}[X,Y]$ verwendet wird (oder $\mathrm{P}^*[X,Y]$, wenn auf absolute Häufigkeiten Bezug genommen werden soll), kann also durch folgendes Schema verdeutlicht werden:

$$\mathrm{P}[X,Y] : \mathcal{P}(\tilde{\mathcal{X}} \times \tilde{\mathcal{Y}}) \longrightarrow [0,1]$$

Ist M irgendeine Merkmalsmenge, d.h. eine Teilmenge von $\tilde{\mathcal{X}} \times \tilde{\mathcal{Y}}$ bzw. ein Element der Potenzmenge von $\tilde{\mathcal{X}} \times \tilde{\mathcal{Y}}$, wird ihr durch die Häufigkeitsfunktion eine Zahl $\mathrm{P}[X,Y](M)$ im Intervall von 0 bis 1 zugeordnet, die den Anteil der Mitglieder von Ω angibt, die Merkmalswerte in M aufweisen:

$$\mathrm{P}[X,Y](M) = \frac{|\{\omega \in \Omega \,|\, (X,Y)(\omega) \in M\}|}{|\,\Omega\,|}$$

Als Beispiel sei etwa $M := \{1\} \times \{20, \ldots, 25\}$. Dann ist $\mathrm{P}[X,Y](M) = 2/10$, d.h. 20 % der Personen in der Referenzmenge Ω sind weiblich und 20 bis 25 Jahre alt.

Art „Für die Mitglieder der Gesamtheit Ω gilt …“ kann bedeuten:

(1) Für jedes Mitglied aus Ω gilt …; oder

(2) Für die Gesamtheit der Mitglieder aus Ω, also für Ω gilt …

Statistische Aussagen, die vom Begriff einer statistischen Verteilung ausgehen, sind stets vom Typ (2), nicht vom Typ (1).

Natürlich müssen zunächst Daten über individuelle Mitglieder einer Gesamtheit erhoben werden, bevor eine statistische Verteilung gebildet werden kann. Insofern bezieht sich die Erhebung statistischer Daten auf individuelle Objekte. Ein Perspektivenwechsel findet jedoch statt, sobald man statistische Verteilungen betrachtet. Die Aufmerksamkeit richtet sich dann auf die Gesamtheit, nicht mehr auf ihre individuellen Mitglieder, anhand derer die Daten gewonnen worden sind. Diese der statistischen Methode eigentümliche Abstraktion wurde vom *International Statistical Institute* (1986: 238) in einer „Declaration of Professional Ethics“ folgendermaßen formuliert:

„Statistical data are unconcerned with individual identities. They are collected to answer questions such as ‘how many?’ or ‘what proportions?’, not ‘who?’. The identities and records of co-operating (or non-cooperating) subjects should therefore be kept confidential, whether or not confidentiality has been explicitly pledged.“

1.2 Statistische Strukturbegriffe

1. Statistische Strukturen und Sachverhalte. Von „Strukturen“ wird in unterschiedlichen Bedeutungen gesprochen. Ein in der statistisch orientierten Literatur verbreiteter Sprachgebrauch verwendet den Strukturbegriff im wesentlichen gleichbedeutend mit dem Begriff einer statistischen Häufigkeitsverteilung. Ich nenne dies den *statistischen Strukturbegriff.*[14] So wird z.B. von einer Altersstruktur der Bevölkerung oder von einer Berufsstruktur der Erwerbstätigen gesprochen.[15] Ulrich Mueller hat diesen statistischen Strukturbegriff folgendermaßen erläutert:

„Die Struktur einer bestimmten Bevölkerung wird beschrieben durch die absolute Zahl der Einheiten sowie die Verteilung der jeweils interessierenden Merkmalsausprägungen bei den Einheiten dieser Bevölkerung zu einem bestimmten Zeitpunkt t.“ (Mueller 1993: 2)

Diese Formulierung stammt aus einer Einführung in die Bevölkerungsstatistik. Aber der statistische Strukturbegriff ist nicht nur in der Demographie

[14]Es sollte beachtet werden, dass auch innerhalb der statistischen Literatur noch in anderen Bedeutungen von „Struktur“ gesprochen wird.

[15]Zahlreiche Illustrationen findet man u.a. bei Bodzenta (1979) und Edinger (1998: 7ff.).

verbreitet,[16] sondern spielt in den meisten Varianten der Sozialforschung eine wichtige Rolle. Dies gilt insbesondere für diejenigen Ansätze, die eine wichtige Aufgabe darin sehen, die Mitglieder einer Gesellschaft in Klassen oder Schichten einzuteilen. Typischerweise meint der Sozialstrukturbegriff dann eine statistische Verteilung der Bevölkerung auf die zuvor konstruierten Klassen bzw. Schichten. Das ist von einigen Autoren als „oberflächlich“ kritisiert worden,[17] hier interessiert uns jedoch der Begriff selbst und erst im Anschluss auch die Frage, was mit ihm erreicht werden kann.

Noch eine terminologische Bemerkung: Da wir den statistischen Strukturbegriff synonym mit dem Begriff einer statistischen Verteilung verwenden, können Feststellungen statistischer Strukturen auch als statistische Sachverhalte bezeichnet werden. Allgemein verstehen wir unter *statistischen Sachverhalten* Feststellungen, die sich sowohl auf statistische Verteilungen als auch auf aus diesen abgeleitete Charakterisierungen (wie z.B. Mittelwerte, Streuungen, Anteilswerte, Raten, Regressionsfunktionen und -koeffizienten) beziehen können.

2. Einige Besonderheiten des statistischen Strukturbegriffs. Es sollte beachtet werden, dass einige mit dem Strukturbegriff oft verbundene Vorstellungen nicht zu seiner statistischen Verwendung passen. Dies betrifft zunächst die Annahme, dass mit dem Strukturbegriff stets auf Beziehungen zwischen irgendwelchen Elementen verwiesen werden soll. George C. Homans (1976: 54) hat das so formuliert:

„[M]any sociologists use “social structure” to refer to some kind of social whole, which can be divided, at least conceptually, into parts, and in which the parts are in some way interdependent, at least in the sense that a change in some of them will be associated with changes in some of the others.“

In dieser Formulierung erinnert Homans an einen relationalen Strukturbegriff. Mit einigen Varianten solcher Vorstellungen beschäftigen wir uns in

[16] Hier noch eine Formulierung des Demographen R. Pressat (1972: 1): „Demography is the discipline that seeks a statistical description of human populations with respect to (1) their demographic structure (the number of the population; its composition by sex, age and marital status; statistics of families, and so on) at a given date, and (2) the demographic events (births, deaths, marriages and terminations of marriages) that take place in them.“ Mit „demographischen Strukturen“ sind hier offenbar Varianten statistischer Strukturen gemeint.

[17] Eine durchaus typische Variante dieser Kritik kommt etwa in folgenden Bemerkungen von Friedrich Fürstenberg (1966: 443) zum Ausdruck: „Es gibt eine Reihe von Autoren, die „Sozialstruktur“ als *statistisches Klassifikationssystem* interpretieren. Sie setzen damit den „Gliederungsaspekt“ des Begriffes, auf den *Karl Martin Bolte* hingewiesen hat, absolut. Autoren mit dieser Blickrichtung stehen häufig der wirtschaftsstatistischen Sichtweise nahe und haben als reine Empiriker ein unreflektiertes Verhältnis zur soziologischen Theorie.“ Man vgl. auch die Kritik von René König (1958b: 259) an „reinen Inventarisierungen einer Bevölkerung“. – Von Vorwürfen dieser Art sind Überlegungen zu unterscheiden, die gegen statistische Strukturbegriffe einwenden, dass soziale Beziehungen unberücksichtigt bleiben. Darauf wird in Abschnitt 2.2 näher eingegangen.

Kapitel 2. Hier soll zunächst darauf hingewiesen werden, dass der statistische Strukturbegriff vollständig unabhängig von irgendwelchen Vorstellungen über Beziehungen konzipiert ist.[18]

Eine zweite Differenz betrifft die Annahme, dass mit dem Strukturbegriff auf Sachverhalte verwiesen wird, die besonders dauerhaft sind oder jedenfalls dauerhafter als Vorgänge oder Prozesse, die sich im Rahmen gegebener Strukturen abspielen. Auch auf diese Konnotation des Strukturbegriffs wird von Homans (1976: 54) hingewiesen. Stefan Hradil (1987: 14) bemerkt dazu: „Es wird, wie immer, wenn der Strukturbegriff Anwendung findet, eine relativ beständige Anordnung von Elementen angesprochen.“[19] Wiederum muss jedoch beachtet werden, dass der statistische Strukturbegriff diese Vorstellung nicht beinhaltet. Die Begriffsbildung hat keinerlei Implikationen für die Frage, wie sich eine statistische Verteilung im Zeitablauf entwickelt. Man kann deshalb auch ganz unproblematisch die Frage stellen, wie sich statistische Strukturen (z.B. Haushaltsstrukturen) im Zeitablauf verändern.

Schließlich muss beachtet werden, dass statistisch definierte Strukturen, also Häufigkeitsverteilungen, nicht mit der Vorstellung eines „Musters“ in Verbindung gebracht werden können. Zwar kann man sinnvoll von der *Form einer statistischen Verteilung* sprechen (insbesondere dann, wenn ein quantitativer Merkmalsraum gegeben ist, so dass es eine lineare Ordnung der Merkmalswerte gibt); bekanntlich können Verteilungen graphisch dargestellt werden und liefern dadurch eine direkte Anschauung ihrer Form, die auch durch statistische Kennzahlen charakterisiert werden kann. Aber solche graphischen Darstellungen vermitteln nicht die Vorstellung eines mit Regelmäßigkeiten assoziierbaren Musters.

3. Unterschiedliche Sozialstrukturbegriffe. Der statistische Strukturbegriff ist so allgemein, dass in zahlreichen Varianten – oder vielleicht besser: Aspekten – von „Sozialstruktur“ gesprochen werden kann. Unterschiede kann es sowohl in den Arten der Objekte geben, auf die man sich bezieht (etwa Personen, Haushalte, Unternehmen oder Regionen), als auch bei den Merkmalsräumen, die zur Charakterisierung der Objekte verwendet werden (etwa Alter, Einkommen und Bildung bei Personen oder Beschäftigtenzahl, Umsatz und Wirtschaftszweig bei Unternehmen).

[18] Leider wird diese Unterscheidung nicht immer beachtet. Zum Beispiel verwendet P. M. Blau, mit dessen Ansatz wir uns weiter unten genauer beschäftigen werden, einen statistischen Sozialstrukturbegriff; scheinbar darauf Bezug nehmend, gibt es jedoch immer wieder Formulierungen, die eigentlich einen relationalen Strukturbegriff voraussetzen (z.B. Blau 1977: 26ff.).

[19] So auch Scheuch und Kutsch (1975: 215): „Struktur bezeichnet das Dauerhafte an einem Gefüge von Elementen.“ Hier setzt auch ein leicht irreführender Kontrast zum Prozessbegriff an, z.B. in einer Formulierung von J. M. Blaut (1971: 19): „The relatively static events are often referred to as ‘structure’; the relative mobile ones as ‘process’ or ‘function’.“

Eine weitere Unterscheidung kann im Anschluss an folgende Ausführungen von Wolfgang Zapf erläutert werden:

„Unter *Sozialstruktur* kann man mindestens dreierlei verstehen. Erstens die demographische Grundgliederung der Bevölkerung und die Verteilung zentraler Ressourcen wie Bildung, Beruf und Einkommen. [. . .]
Zweitens kann man unter Sozialstruktur – unter Einschluß von Werten und Mentalitäten – die Zusammenfassung dieser Gliederungen in soziale Klassen und Schichten verstehen [. . .].
Drittens gibt es den anspruchsvolleren Begriff von Sozialstruktur als dem jeweils historisch ausgeprägten System gesellschaftlicher Ordnungen oder Grundinstitutionen [. . .].“ (Zapf 1995: 187)

Offenbar handelt es sich in den ersten beiden Fällen um Varianten des statistischen Strukturbegriffs (erst Zapfs dritte Variante führt zu einem grundsätzlich anderen Zugang zur Idee einer Sozialstruktur). In der ersten Variante sind statistische Verteilungen gemeint, deren Merkmalsräume unmittelbar auf in der gesellschaftlichen Praxis übliche Unterscheidungen verweisen; in der zweiten Variante setzt der Sozialstrukturbegriff die vorgängige Konstruktion eines Schemas zur Klassifikation der Mitglieder einer Gesellschaft voraus. Beide Varianten werden in den folgenden Paragraphen etwas näher besprochen.

4. Der Sozialstrukturbegriff bei Peter M. Blau. Der ersten Variante lassen sich die meisten statistischen Beiträge zur Beschreibung gesellschaftlicher Verhältnisse zurechnen. Wegen seines theoretischen Anspruchs ist in diesem Zusammenhang besonders der Forschungsansatz von Peter M. Blau, von ihm selbst als „makro-strukturell“ bezeichnet, von Interesse:

„Macrostructural concepts refer to people’s distribution in various dimensions and the degrees to which these dimensions of social differences among people are related. Macrosociology is concerned primarily with large populations – composed of many thousands or even millions of persons. My endeavor is to develop a systematic theoretical scheme for the study of macrostructures and their impact on social life.“ (Blau 1994a: 1)

Bei den „Dimensionen“ kann es sich um beliebige Attribute (Merkmalsräume) handeln, durch die sich die Mitglieder einer Gesellschaft unterscheiden lassen.[20] Wichtig ist der Hinweis, dass Beziehungen zwischen diesen „Dimensionen“ ermittelt werden sollen, denn das setzt voraus, dass man sie nicht sogleich zu Klassifizierungen zusammenfasst. Dem entspricht folgende Definition:

„Social structure can be conceptualized as a multidimensional space of social positions among which a population is distributed.“ (Blau 1994a: 4)

[20] Blau spricht in diesem Zusammenhang von „sozialen Positionen“, meint aber „any difference among people in terms of which they make social distinctions among themselves in social intercourse“ (Blau 1994a: 3).

Diese Begriffsbildung kann gut mithilfe einer mehrdimensionalen statistischen Variablen ausgedrückt werden:

$$(X_1, \ldots, X_m) : \Omega \longrightarrow \tilde{\mathcal{X}}_1 \times \cdots \times \tilde{\mathcal{X}}_m$$

In dieser Formulierung bezieht sich Ω auf eine Population (eine irgendwie abgegrenzte Gesellschaft), und die Komponenten $X_1, \ldots, X_m$ der mehrdimensionalen Variablen erfassen Eigenschaften der Mitglieder von Ω in den m „Dimensionen“ (Merkmalsräumen) $\tilde{\mathcal{X}}_1, \ldots, \tilde{\mathcal{X}}_m$. Was Blau „Sozialstruktur“ nennt, entspricht formal der Verteilung dieser m-dimensionalen statistischen Variablen. Anhand dieser Formulierung wird auch deutlich, was damit gemeint ist, Beziehungen zwischen den „Dimensionen“ der Sozialstruktur zu ermitteln. Es geht um Charakterisierungen gemeinsamer Verteilungen der Komponenten von $(X_1, \ldots, X_m)$, etwa durch Korrelationen (Blau 1994a: 5) oder allgemeiner durch Regressionsfunktionen.[21]

5. Bezugseinheiten statistisch definierter Sozialstrukturen. Blaus Sozialstrukturbegriff bezieht sich zunächst auf Populationen, also irgendwie abgegrenzte Gesamtheiten von Menschen. Vollständig analog kann man aber auch bei anderen Bezugseinheiten Aussagen über statistische Strukturen (Merkmalsverteilungen) machen; insbesondere kann man sich auf Haushalte, Unternehmen und Regionen beziehen.

In diesem Zusammenhang sollte auch auf eine Ambivalenz im Sprachgebrauch geachtet werden, die sich anhand der folgenden Bemerkung von Blau (1974: 615f.) erläutern lässt:

„The concept of social structure is used widely in sociology, often broadly, and with a variety of meanings. [. . .] A generic difference is whether social structure is conceived explicitly as being composed of different elements and their interrelations or abstractly as a theoretical construct or model. [. . . [22]] If one adopts the first view, as I do, that social structure refers to the differentiated interrelated parts in a collectivity, not to theories about them, the fundamental question is how these parts and their connections are conceived.
My concept of social structure starts with simple and concrete definitions of the component parts and their relations. The parts are groups or classes of people, such as men and women, ethnic groups, or socioeconomic strata; more precisely, they are the positions of people in different groups and strata. The connections among as well as within the parts are the social relations of people that find expression in their social interaction and communication.“

Zunächst erscheint diese Aussage als ein Widerspruch zur oben in § 4 zitierten Bezugnahme auf Populationen, also Gesamtheiten von Individuen. Der Widerspruch verschwindet jedoch, wenn man zwischen Individuen

[21] Regressionsfunktionen werden in Abschnitt 1.3 besprochen.

[22] Hier erwähnt Blau kurz die Auseinandersetzung zwischen Radcliffe-Brown und Lévi-Strauss über den theoretischen Status des (Sozial-) Strukturbegriffs. Eine ausführliche Diskussion findet man bei Michael Oppitz (1975: 33ff.).

(oder allgemeiner irgendwelchen Bezugseinheiten) und Positionen (im Sinne Blaus) unterscheidet. Unser Begriff einer statistischen Variablen macht das deutlich. Denkt man an das Schema $X : \Omega \longrightarrow \tilde{\mathcal{X}}$, repräsentiert Ω die Population, und die Elemente des Merkmalsraums $\tilde{\mathcal{X}}$ sind die Positionen, die zur Charakterisierung der Mitglieder von Ω verwendet werden sollen. Die statistische Variable X induziert nun außerdem eine Partition der Population: Jeder Position $x \in \tilde{\mathcal{X}}$ entspricht eine Menge

$$X^{-1}(\{x\}) = \{\omega \in \Omega \,|\, X(\omega) = x\}$$

die aus denjenigen Mitgliedern von Ω besteht, die die Position x haben. Zwar muss man, wie Blau selbst bemerkt, zwischen der Position x und der Teilpopulation $X^{-1}(\{x\})$ unterscheiden; aber die Häufigkeit der Position x in der statistischen Struktur entspricht natürlich dem Umfang der korrespondierenden Teilpopulation.

Eine weitere Quelle für Unklarheiten ist allerdings Blaus Bemerkung, dass sich sein statistischer Strukturbegriff auf „component parts *and their relations*“ bezieht. Wie bereits in § 2 bemerkt worden ist, passt diese Rhetorik nicht zur Verwendung eines statistischen Strukturbegriffs. Dass zwei oder mehr Personen irgendeine statistisch erfassbare Eigenschaft gemeinsam haben, kann zwar möglicherweise zur Feststellung von Ähnlichkeiten dienen, begründet aber keine substantielle Beziehung zwischen diesen Personen.[23] Umgekehrt verweist Blaus Rede von „social interaction and communication“ zwar auf substantielle Beziehungen (und also indirekt auf einen relationalen Sozialstrukturbegriff), auf *diese* Beziehungen nimmt aber Blaus statistischer Sozialstrukturbegriff gar keinen Bezug.

6. Wie entstehen statistische Sachverhalte? Dieser Frage können unterschiedliche Bedeutungen gegeben werden. Zur Erläuterung betrachten wir folgendes Bild:

○ ○ ● ○ ● ● ○ ○ ○ ●

Das Bild stellt einen Sachverhalt dar, der aus 10 Objekten besteht, von denen 4 schwarz, die übrigen nicht schwarz sind. Offenbar handelt es sich in dieser Darstellung nicht bereits um einen statistischen Sachverhalt. Ein statistischer Sachverhalt *entsteht erst durch eine spezifische Konzeptualisierung*, die drei wesentliche Schritte umfasst: (a) die Konzeption einer statistischen Variablen; (b) einen (realen oder fiktiven) *datenerzeugenden Prozess*, der Informationen über die Werte der Variablen liefert; und (c) rechnerische Operationen, die es erlauben, sich gedanklich auf die Verteilung der statistischen Variablen zu beziehen.

In unserem Beispiel können diese drei Schritte offenbar problemlos ausgeführt werden. Zunächst kann man die 10 Objekte durch eine Objektmenge $\Omega := \{\omega_1, \ldots, \omega_{10}\}$ repräsentieren (wobei die Zuordnung der Namen

[23] Eine ausführliche Darlegung dieser Kritik findet man bei Bates und Peacock (1989).

beliebig erfolgen kann); und es kann ein Merkmalsraum $\tilde{\mathcal{Y}} := \{0, 1\}$ festgelegt werden (wobei etwa 1 für schwarz und 0 für nicht schwarz steht), so dass schließlich der erste Schritt durch die Definition einer statistischen Variablen $Y : \Omega \longrightarrow \tilde{\mathcal{Y}}$ abgeschlossen werden kann, durch die jedem der 10 Objekte ein Wert im Merkmalsraum $\tilde{\mathcal{Y}}$ zugeordnet wird. In einem zweiten Schritt können dann Daten erzeugt werden. Das ist in diesem Beispiel direkt und vollständig möglich und liefert eine Tabelle:

ω	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8	ω_9	ω_{10}
$Y(\omega)$	0	0	1	0	1	1	0	0	0	1

Diese Tabelle enthält das Datenmaterial für den dritten Schritt, in dem die Verteilung der Variablen Y berechnet wird. Auch dieser Schritt ist in diesem Beispiel direkt durchführbar und erfordert nicht einmal eine tabellarische oder graphische Darstellung; es genügt die Angabe, dass 40 % der Objekte schwarz, die übrigen nicht schwarz sind.

Diese Aussage beschreibt nun einen statistischen Sachverhalt, und somit illustriert das Beispiel *eine* Antwort auf unsere Ausgangsfrage: Ein statistischer Sachverhalt entsteht durch eine gedankliche und praktische Konstruktion, die in den drei genannten Schritten abläuft. Diese Antwort macht deutlich, dass statistische Sachverhalte durch spezifische gedankliche und praktische Konstruktionen entstehen.

7. Datenerzeugende und substantielle Prozesse. Außerdem macht die Antwort deutlich, dass die Konstruktion eines statistischen Sachverhalts voraussetzt, dass die Mikro-Sachverhalte, auf die sich ein datenerzeugender Prozess beziehen kann, bereits existieren. Infolgedessen können zwei Arten von Fragen unterschieden werden:

- Wie entstehen diese Mikro-Sachverhalte bzw. wie sind sie entstanden? (Wie sind z.B. die schwarzen und nicht-schwarzen Objekte in dem oben in § 6 angeführten Bild entstanden?)
- Wie entsteht ausgehend von bereits bestehenden Mikro-Sachverhalten ein statistischer Sachverhalt?

Das folgende Bild veranschaulicht die Unterscheidung:

$$\left.\begin{array}{l} \dashrightarrow\; Y(\omega_1) \\ \qquad\vdots \\ \dashrightarrow\; Y(\omega_n) \end{array}\right\} \Longrightarrow \mathrm{P}[Y]$$

Durch $Y(\omega_1), \ldots, Y(\omega_n)$ werden die Mikro-Sachverhalte angedeutet,[24] wobei jedem dieser Mikro-Sachverhalte ein Prozesspfeil $\dashrightarrow$ zugeordnet ist,

[24] Dies ist offenbar eine verkürzte Darstellung, da es sich bei $Y(\omega_1), \ldots, Y(\omega_n)$ nicht um Sachverhalte, sondern um Elemente des Merkmalsraums von Y handelt. Eine explizite Notation für die korrespondierenden Mikro-Sachverhalte findet man bei Rohwer und Pötter (2002b: 211ff.).

der einen *substantiellen Prozess* andeuten soll, durch den der jeweilige Mikro-Sachverhalt entstanden ist. Andererseits wird durch den Pfeil $\Longrightarrow$ auf den *Konstruktionsprozess* verwiesen, durch den der statistische Sachverhalt $\mathrm{P}[Y]$ entsteht, wobei – wie das Bild zeigt – die Mikro-Sachverhalte vorausgesetzt werden. Es müssen also insgesamt drei Arten von Prozessen unterschieden werden:

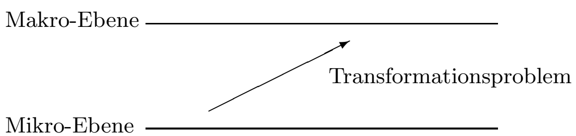
– *Substantielle Prozesse,* durch die den Elementen einer Objektmenge individuell zurechenbare Mikro-Sachverhalte entstehen;

– *datenerzeugende Prozesse,* durch die Informationen (die Daten) über die für die Datenerzeugung vorauszusetzenden Mikro-Sachverhalte entstehen;[25] und

– *Rechenprozesse,* durch die ausgehend von jeweils gegebenen Mengen von Daten (die als Resultat eines datenerzeugenden Prozesses entstanden sind) statistische Sachverhalte konstruiert werden.

Offenbar interessieren in erster Linie die substantiellen Prozesse, durch die die Mikro-Sachverhalte in der sozialen Realität entstehen; man möchte verstehen, wie diese Prozesse ablaufen und wodurch sie bedingt werden.

8. Statistische Sachverhalte im Mikro-Makro-Schema. Eine wichtige Frage betrifft die theoretische Konzeptualisierung substantieller Prozesse. Hier soll zunächst darauf hingewiesen werden, dass eine Einordnung statistischer Sachverhalte in ein Mikro-Makro-Schema dafür kaum hilfreich ist. Zur Erläuterung kann folgendes Bild dienen:



In diesem Bild gibt es eine Mikro-Ebene, mit der auf individuelle Menschen, ihre Tätigkeiten und ihnen zurechenbare Mikro-Sachverhalte verwiesen werden soll, und eine Makro-Ebene, die sich zunächst allgemein auf „kollektive Phänomene“ bezieht. Außerdem gibt es einen Pfeil, der von der Mikro- zur Makro-Ebene führt und die Vorstellung andeuten soll, dass die „kollektiven Phänomene“ der Makro-Ebene irgendwie aus auf der Mikro-Ebene fixierbaren Sachverhalten und Vorgängen resultieren.[26] Die

[25] Es sei angemerkt, dass der Begriff eines datenerzeugenden Prozesses in der Literatur gelegentlich auch anders verwendet wird, nämlich als Verweis auf die unter (a) genannten substantiellen Prozesse; man vgl. dazu die Hinweise und Literaturangaben bei Rohwer und Pötter (2002b: 19ff.).

[26] In einem vollständigen Mikro-Makro-Schema gibt es auch einen Pfeil, der von der Makro- zur Mikro-Ebene führt und die Vorstellung andeuten soll, dass die auf der Mikro-Ebene fixierbaren Sachverhalte und Vorgänge auch von Sachverhalten auf der Makro-Ebene abhängig sind. Davon kann hier jedoch abgesehen werden.

Frage, wie dies geschieht, wird im Anschluss an S. Lindenberg (1977) oft als *Transformationsproblem* bezeichnet.

Zu überlegen ist, welche Arten von Sachverhalten auf der Makro-Ebene dieses Schemas verortet werden können. In der Literatur findet man Hinweise auf sehr unterschiedliche Arten von Sachverhalten. So spricht etwa Lindenberg (1977: 49) von „kollektiven Phänomenen“ „wie z.B. kollektive Handlungen (etwa Streiks), Strukturen (etwa Statusstrukturen), Verteilungen (etwa Einkommensverteilungen), Institutionen (etwa Institutionalisierung von Konflikten)“. Offenbar wird hier einerseits auf statistische Sachverhalte Bezug genommen,[27] andererseits aber auch auf zahlreiche andere Arten „kollektiver Phänomene“. Infolgedessen muss jedoch auch der Pfeil, der von der Mikro- zur Makro-Ebene führt, jeweils unterschiedlich interpretiert werden. Handelt es sich z.B. um einen Streik, kann man sinnvoll von einem Sachverhalt sprechen, der aus den Tätigkeiten einer Mehrzahl beteiligter Akteure resultiert. Handelt es sich dagegen um statistische Sachverhalte, kann man dies nicht sagen. Man kann zwar in vielen Fällen sinnvoll davon sprechen, dass die dem statistischen Sachverhalt korrespondierenden Mikro-Sachverhalte aus Tätigkeiten von Akteuren resultieren. Sowohl diese Tätigkeiten wie auch die aus ihnen resultierenden Mikro-Sachverhalte gehören jedoch zur Mikro-Ebene des Schemas. Wenn man also statistische Sachverhalte auf der Makro-Ebene ansiedelt, entspricht dem Pfeil, der von der Mikro- zur Makro-Ebene führt, auch kein substantieller Prozess, sondern nur der statistische Konstruktionsprozess, der *aus bereits entstandenen* Mikro-Sachverhalten eine spezifische Art ihrer Beschreibung erzeugt.

1.3 Statistische Regressionsrechnung

Ein primärer Zweck statistischer Methoden besteht darin, Daten, die sich zunächst auf die Elemente von Gesamtheiten beziehen, so darzustellen, dass man Informationen über statistische Eigenschaften der Gesamtheiten gewinnt. Eine darüber hinausgehende Frage ist, wie statistische Begriffsbildungen und Methoden zur Ermittlung von Bedingungsverhältnissen verwendet werden können. Oft werden Methoden der Regressionsrechnung verwendet. In diesem Abschnitt wird der konzeptionelle Ansatz besprochen, wobei von statistischen Variablen ausgegangen wird; später (insbesondere in Kapitel 5) werden auch stochastische Regressionsmodelle behandelt.

1. Definition bedingter Verteilungen. Grundlegend für die meisten analyti-

[27] Ebenso sprechen Büschges, Abraham und Funk (1998: 18) von einer Makro-Ebene, „die kollektive Phänomene wie statistische Verteilungen (z.B. die Quote der Frauenerwerbsbeteiligung in einer Gesellschaft) oder kollektives Verhalten (wie Demonstrationen vieler Individuen) abbildet.“

schen Verwendungen statistischer Methoden ist der Begriff einer *bedingten Verteilung*. Wir verwenden allgemein die Notation

$$\mathrm{P}[\text{Variablen} \,|\, \text{Bedingungen}]$$

womit die Verteilung derjenigen Variablen gemeint ist, die vor dem Bedingungsstrich | stehen, und zwar eingeschränkt auf diejenige Teilgesamtheit der Referenzmenge Ω, bei der die hinter dem Bedingungsstrich genannten Bedingungen erfüllt sind. Bedingte Verteilungen sind also, wie alle Häufigkeitsverteilungen, Funktionen; bestimmte numerische Werte (Häufigkeiten) erhält man erst, wenn man sich auf den Wert der Funktion bei einem bestimmten Argument bezieht (das in runden Klammern an den Funktionsnamen angehängt wird).

Zur Illustration beziehen wir uns auf die im Abschnitt 1.1 verwendete zweidimensionale Variable $(X,Y) : \Omega \longrightarrow \tilde{\mathcal{X}} \times \tilde{\mathcal{Y}}$, bei der X das Geschlecht und Y das Alter von 10 Personen erfasst (die Daten findet man in Tabelle 1.1-1). Einige Beispiele für bedingte Verteilungen, die in diesem Fall gebildet werden können, sind die folgenden:

- $\mathrm{P}[Y|X=0]$. Dies ist die Verteilung der Variablen Y bei derjenigen Teilgesamtheit, bei der die Variable X den Wert 0 hat; es handelt sich also um die Altersverteilung der Männer. Wie gesagt, ist dies eine Funktion. Ein bestimmter Wert wäre etwa $\mathrm{P}[Y|X=0](Y \geq 25) = 0.5$, d.h. 50% der Männer sind mindestens 25 Jahre alt. In analoger Bedeutung ist $\mathrm{P}[Y|X=1]$ die Altersverteilung der Frauen.
- $\mathrm{P}[X|Y \geq 25]$ ist die Verteilung von X bei den Personen, die mindestens 25 Jahre alt sind. Also ist z.B. $\mathrm{P}[X|Y \geq 25](1) = 0.5$, d.h. 50 % der mindestens 25 Jahre alten Personen sind weiblich.
- $\mathrm{P}[Y|Y \geq 25, X=1]$ ist die Altersverteilung der Frauen, die mindestens 25 Jahre alt sind. Dieses Beispiel zeigt auch, dass Variablen gleichzeitig vor und hinter dem Bedingungsstrich verwendet werden können.

Wie bei einfachen (unbedingten) Verteilungen können unterschiedliche Schreibweisen verwendet werden, wenn man sich auf bestimmte Argumente bezieht; beispielsweise $\mathrm{P}(Y \geq 25|X=0)$ anstelle von $\mathrm{P}[Y|X=0](Y \geq 25)$ im ersten Beispiel und $\mathrm{P}(X=1|Y \geq 25)$ anstelle von $\mathrm{P}[X|Y \geq 25](1)$ im zweiten Beispiel.

2. Statistische Regressionsfunktionen. Wenn von statistischer Regressionsrechnung gesprochen wird, sind allgemein Methoden zur Darstellung bedingter Verteilungen gemeint. Ein Verständnis gewinnt man durch eine Unterscheidung allgemeiner und spezieller Regressionsfunktionen.[28] Ausgangspunkt ist eine zweidimensionale Variable $(X,Y) : \Omega \longrightarrow \tilde{\mathcal{X}} \times \tilde{\mathcal{Y}}$ (wobei sowohl X als auch Y mehrdimensional sein können). Dann wird eine

[28] Eine ausführliche Diskussion findet man bei Rohwer und Pötter (2001, Teil II).

der beiden Komponenten als *unabhängige Variable* (auch *Regressorvariable* genannt), die andere als *abhängige Variable* bestimmt; wir wählen für die folgende Darstellung X als unabhängige und Y als abhängige Variable. Schließlich kann eine *allgemeine Regressionsfunktion* definiert werden, die jedem möglichen Merkmalswert $x \in \tilde{\mathcal{X}}$ die bedingte Verteilung $\mathrm{P}[Y|X=x]$ zuordnet.

Der Zweck dieser Konstruktion besteht darin, dass man ermitteln möchte, wie die Verteilung der abhängigen Variablen durch jeweils spezielle Werte der unabhängigen Variablen bedingt wird. Allerdings sind allgemeine Regressionsfunktionen gewissermaßen Funktionen zweiter Ordnung, da es sich bei ihren Werten um bedingte Verteilungen, also wiederum um Funktionen handelt. Folgendes Bild veranschaulicht dies:

$$\underbrace{x \longrightarrow \overbrace{(y \longrightarrow \mathrm{P}[Y|X=x](y))}^{\text{bedingte Verteilung}}}_{\text{allgemeine Regressionsfunktion}}$$

Eine allgemeine Regressionsfunktion ordnet jedem möglichen Wert x der unabhängigen Variablen X eine bedingte Verteilung zu, die jedoch selbst eine Funktion ist, die jedem möglichen Wert y der abhängigen Variablen Y eine bedingte Häufigkeit $\mathrm{P}[Y|X=x](y)$ zuordnet.

Somit entsteht die Frage, wie man Regressionsfunktionen darstellen kann. Die meisten Vorschläge folgen einem einfachen Grundgedanken, der darin besteht, als Werte einer Regressionsfunktion nicht bedingte Verteilungen im Sinne von Funktionen zu verwenden, sondern Zahlen, durch die die bedingten Verteilungen charakterisiert werden können. Regressionsfunktionen nehmen dann folgende Form an:

$$x \longrightarrow \text{Charakterisierung von } \mathrm{P}[Y|X=x]$$

Dadurch werden Regressionsfunktionen zu gewöhnlichen Funktionen; wir nennen sie *spezielle Regressionsfunktionen*. Durch eine spezielle Regressionsfunktion wird jedem Merkmalswert $x \in \tilde{\mathcal{X}}$ eine bestimmte Zahl zugeordnet, die die bedingte Verteilung $\mathrm{P}[Y|X=x]$ charakterisiert.

Offenbar können spezielle Regressionsfunktionen auf viele unterschiedliche Weisen definiert werden. Hauptsächlich werden jedoch die folgenden drei Charakterisierungen verwendet:

- Bedingte Mittelwerte; spezielle Regressionsfunktionen sehen dann folgendermaßen aus: $x \longrightarrow \mathrm{M}(Y|X=x)$. Jedem Merkmalswert $x \in \tilde{\mathcal{X}}$ wird der durch ihn bedingte Mittelwert der abhängigen Variablen zugeordnet.
- Bedingte Quantile; spezielle Regressionsfunktionen haben dann die Gestalt: $x \longrightarrow \mathrm{Q}_p(Y|X=x)$. Jedem Merkmalswert $x \in \tilde{\mathcal{X}}$ werden ein oder mehrere durch ihn bedingte Quantile der Verteilung der abhängi-

gen Variablen zugeordnet.[29]

- Bedingte Häufigkeiten; spezielle Regressionsfunktionen haben in diesem Fall die Form: $x \longrightarrow \mathrm{P}[Y|X=x](y)$. Hier gibt es für jeden Merkmalswert $y \in \tilde{\mathcal{Y}}$ eine spezielle Regressionsfunktion, die jedem Merkmalswert $x \in \tilde{\mathcal{X}}$ die durch ihn bedingte Häufigkeit für den Merkmalswert $y \in \tilde{\mathcal{Y}}$ zuordnet.

Sowohl bei allgemeinen als auch bei speziellen Regressionsfunktionen handelt es sich um statistische Sachverhalte, so wie dieser Begriff in Abschnitt 1.2 (§ 1) definiert worden ist. Dem entspricht die Bemerkung von V. R. McKim (1997: 7), „that regression is providing new facts, not interpretations or explanations of facts.“ Man kann hinzufügen, dass es sich bei diesen „new facts“ um *aus Daten konstruierte* Sachverhalte handelt.

3. Beispiel: Autofahrer an einer Ampel. Um die Verwendung von Regressionsfunktionen zu verdeutlichen, beginnen wir mit einem einfachen Beispiel: Autofahrer, die sich einer Straßenkreuzung nähern, an der es eine Ampel gibt. Angenommen, wir beobachten für einen gewissen Zeitraum die Straßenkreuzung. Die Beobachtungen können als Werte einer zweidimensionalen Variablen $(X,Y): \Omega \longrightarrow \tilde{\mathcal{X}} \times \tilde{\mathcal{Y}}$ repräsentiert werden. Die Elemente von Ω sind die Situationen, in denen sich ein Autofahrer der Ampel nähert. X erfasst, ob die Ampel rot (1) oder nicht rot (0) ist; Y erfasst, ob das Auto anhält (1) oder nicht anhält (0). Insgesamt sind 100 Situationen beobachtet und folgende Werte festgestellt worden:

x	y	Anzahl
0	0	47
0	1	0
1	0	3
1	1	50

D.h. bei ‘rot’ haben 50 Autos angehalten, 3 sind jedoch weitergefahren; und bei ‘nicht rot’ sind die Autos ausnahmslos weitergefahren.

In diesem Beispiel sind die Daten so überschaubar, dass es nicht erforderlich ist, Regressionsfunktionen zu berechnen; trotzdem kann man es tun. Man muss sich dann zunächst entscheiden, welche der beiden Variablen als unabhängig und welche als abhängig betrachtet werden soll. Trotz der formalen Symmetrie liegt es in diesem Beispiel natürlich nahe, X als unabhängige und Y als abhängige Variable zu betrachten. Eine allgemeine Regressionsfunktion ordnet dann jedem möglichen Wert von X (dem Ampelsignal) eine durch ihn bedingte Verteilung von Y (des Verhaltens

[29] Sei Y irgendeine statistische Variable mit der Verteilungsfunktion F. Dann ist das durch $\mathrm{Q}_p(Y)$ bezeichnete p-Quantil der Verteilung von Y eine Zahl, für die näherungsweise gilt: $F(Q_p(Y)) \approx p$. Dementsprechend ist $\mathrm{Q}_p(Y|X=x)$ das p-Quantil der durch $X=x$ bedingten Verteilung von Y.

der Autofahrer) zu. Stattdessen kann man auch spezielle Regressionsfunktionen bilden, wobei es sich in diesem Beispiel anbietet, Anteilswerte zu verwenden, z.B. den Anteil der Autofahrer, die ihr Auto anhalten.

Bereits an diesem einfachen Beispiel kann man sich auch verdeutlichen, dass Regressionsfunktionen keine Abbildungen in den Merkmalsraum der abhängigen Variablen sind; sie unterscheiden sich also von Funktionen der Form

$$g: \tilde{\mathcal{X}} \longrightarrow \tilde{\mathcal{Y}}$$

Bei diesem Schema liefert die Funktion g für jeden Wert $x \in \tilde{\mathcal{X}}$ einen Wert $g(x)$, der ein Element des Merkmalsraums $\tilde{\mathcal{Y}}$ ist. Offenbar gilt dies nicht für allgemeine Regressionsfunktionen, deren Werte bedingte Verteilungen sind. Aber auch die Werte spezieller Regressionsfunktionen können im Allgemeinen nicht sinnvoll als Werte des Merkmalsraums der abhängigen Variablen interpretiert werden. Das ist wiederum unmittelbar deutlich, wenn (wie in unserem Beispiel) die bedingten Verteilungen durch Häufigkeiten (Anteilswerte) charakterisiert werden. Ein irreführender Eindruck kann höchstens bei der Mittelwertregression entstehen, bei der jedem Wert der unabhängigen Variablen der Mittelwert der durch ihn bedingten Verteilung der abhängigen Variablen zugeordnet wird. Aber auch wenn es sich dabei um einen bestimmten Wert im Merkmalsraum der abhängigen Variablen handelt (was nicht unbedingt der Fall sein muss), muss ein Mittelwert in seiner Bedeutung (nämlich als Charakterisierung einer statistischen Verteilung) von einem individuell zurechenbaren Merkmalswert unterschieden werden.

4. Beispiel: Ausgaben privater Haushalte. Jetzt betrachten wir ein anderes Beispiel, in dem es sowohl für die unabhängige als auch für die abhängige Variable einen quantitativen Merkmalsraum gibt. Die Variable (X,Y) bezieht sich in diesem Beispiel auf eine Gesamtheit privater Haushalte; die unabhängige Variable X erfasst das in einem bestimmten Monat für Ausgaben verfügbare Einkommen, und die abhängige Variable Y erfasst die in diesem Monat getätigten Ausgaben für Nahrungsmittel (einschließlich Getränke und Tabakwaren). Da es sich um eine quantitative abhängige Variable handelt, bietet sich eine Mittelwertregression $x \longrightarrow \mathrm{M}(Y \mid X=x)$ an, durch die jedem möglichen Wert x der unabhängigen Variablen der Mittelwert der Nahrungsausgaben der Haushalte mit dem verfügbaren Einkommen x zugeordnet wird.

Für die Berechnung dieser Regressionsfunktion wären allerdings Individualdaten erforderlich, die uns nicht zur Verfügung stehen. Stattdessen verwenden wir die in Tabelle 1.3-1 angegebenen Daten, die aus der Einkommens- und Verbrauchsstichprobe (EVS) des Jahres 1998 stammen. Die Haushalte wurden vom Statistischen Bundesamt nach ihrem monatlichen Haushaltsnettoeinkommen den in der Tabelle angegebenen Einkommensklassen zugeordnet, dann wurden für jede Einkommensklasse das

Tabelle 1.3-1 Ausgaben privater Haushalte für Nahrungsmittel (einschl. Getränke und Tabakwaren), differenziert nach dem klassifizierten monatlichen Haushaltsnettoeinkommen in DM. Angaben nach der Einkommens- und Verbrauchsstichprobe 1998. Quelle: Statistisches Jahrbuch 2001: 573.

Einkommensklasse von ... bis unter	Ausgabefähiges Einkommen	Ausgaben für Nahrungsmittel in DM	in %
unter 1800	1383	269	19.45
1800 – 2500	2196	341	15.53
2500 – 3000	2788	391	14.02
3000 – 4000	3543	473	13.35
4000 – 5000	4566	584	12.79
5000 – 7000	6057	677	11.18
7000 – 10000	8422	775	9.20
10000 – 35000	13843	894	6.46

durchschnittliche für Ausgaben verfügbare Einkommen und der Durchschnittswert der Ausgaben für Nahrungsmittel berechnet. Somit wird durch diese Rechnungen bereits eine Mittelwertregression durchgeführt: den durchschnittlichen verfügbaren Einkommen werden durchschnittliche Ausgaben für Nahrungsmittel zugeordnet. Natürlich kann man stattdessen auch die durchschnittlichen Anteile der Ausgaben für Nahrungsmittel als eine Funktion des verfügbaren Einkommens betrachten; Abb. 1.3-1 zeigt dies in Form einer graphischen Darstellung. Durch die Daten der Tabelle 1.3-1 sind zwar nur die als Punkte eingezeichneten Funktionswerte begründbar. Man kann aber vermuten, dass eine nicht-parametrische Regression mit den zugrundeliegenden Individualdaten einen zur gestrichelt eingezeichneten Linie sehr ähnlichen Funktionsverlauf liefern würde.

5. Statistische und substantielle Bedingungen. Schließlich stellt sich die Frage, ob und ggf. wie die bei der Bildung bedingter Verteilungen formal zur Konditionierung verwendeten Variablen auch in irgendeinem substantiellen Sinn als Bedindungen interpretiert werden können. Man sollte sich klarmachen, dass der begriffliche Rahmen der Regressionsrechnung dies weder voraussetzt noch impliziert. Denn bei der Bildung einer bedingten Verteilung $\mathrm{P}[Y \mid \text{Bedingung}]$ bezieht sich die hinter dem Bedingungsstrich angeführte Bedingung nicht auf einen Prozess, sondern spezifiziert nur eine Referenzmenge, nämlich diejenige Teilmenge der Objektmenge Ω, für die die Verteilung von Y ausgewiesen werden soll. Um dagegen bei einer Regressionsrechnung Werte der unabhängigen Variablen als effektive Bedingungen interpretieren zu können, wäre es erforderlich, zunächst einen substantiellen Prozess zu konzipieren, durch den Werte der abhängigen Variablen entstehen – um dann darlegen zu können, wie dieser substantielle Prozess von Werten der unabhängigen Variablen abhängt.

Allerdings setzt die Regressionsrechnung nicht einmal die Existenz eines solchen substantiellen Prozesses voraus, was man schon daraus erkennt,

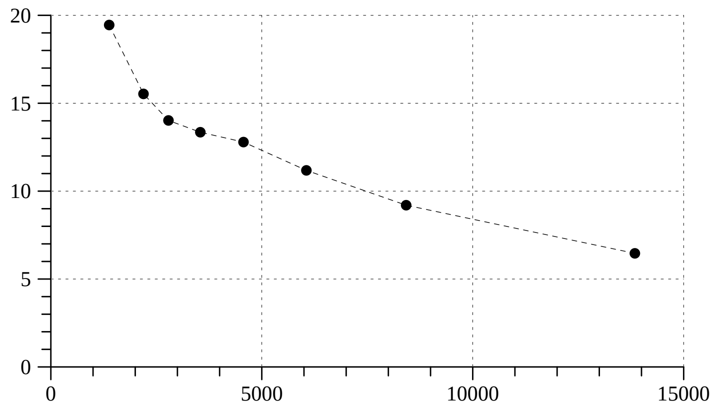


Abb. 1.3-1 Anteile der Ausgaben für Nahrungsmittel (y-Achse: in %) als Funktion des für Ausgaben verfügbaren Haushaltseinkommens (x-Achse: in DM). Daten aus der EVS 1998 (Tabelle 1.3-1).

dass unabhängige und abhängige Variablen beliebig vertauscht werden können. Insbesondere liefert die Regressionsrechnung keine begrifflichen Hilfsmittel zur Repräsentation substantieller Prozesse. Sofern man sich gedanklich auf solche Prozesse beziehen möchte, müssen sie vielmehr jenseits der statistischen Begriffsbildungen und Rechnungen vorstellbar gemacht werden; oder anders formuliert: die Ergebnisse einer Regressionsrechnung müssen im Hinblick auf einen irgendwie vorstellbar gemachten substantiellen Prozess interpretiert werden.

6. Bezugnahmen auf substantielle Prozesse. Betrachten wir das oben angeführte Beispiel, in dem sich Autofahrer einer Ampel nähern. Ob die Autos anhalten oder nicht, wird offenbar nicht durch die Ampelsignale bewirkt, sondern hängt vom Verhalten der Autofahrer ab. Ein Ampelsignal bildet nur einen Aspekt einer Situation, in der sich ein Autofahrer so oder so verhält, und tatsächlich schränkt es nicht einmal seine effektiven Handlungsmöglichkeiten ein. Um zu erklären, wie Ampelsignale gleichwohl als Bedingungen verstanden werden können, muss man sich also in irgendeiner Weise darauf beziehen, wie sie von Autofahrern wahrgenommen und verarbeitet werden.

Ebenso muss in dem zweiten Beispiel auf menschliche Tätigkeiten Bezug genommen werden, um zu verstehen, wie Ausgaben für Nahrungsmittel zustande kommen und u.a. vom jeweils verfügbaren Haushaltseinkommen abhängen. Entsprechend gilt bei vielen Anwendungen der Regressionsrechnung in den Sozialwissenschaften, dass bei der Konzeption substantieller Prozesse auf Akteure Bezug genommen werden muss. In diesen Fällen kann man sich oft an einem Schema der folgenden Art orientieren:

$$\begin{array}{ccc} \boxed{\text{Substantielle Prozesse (durch Akteure)}} & \rightarrow & y \\ \uparrow & & \\ x & & \end{array}$$

An den Prozessen, durch die bestimmte Werte (y) der abhängigen Variablen Y entstehen, sind Akteure beteiligt; und es stellt sich dann die Frage, ob und ggf. wie bestimmte Werte (x) der unabhängigen Variablen X als Bedingungen für Verhaltensweisen und Tätigkeiten dieser Akteure, soweit sie am Zustandekommen von Werten der abhängigen Variablen beteiligt sind, interpretiert werden können.

Kapitel 2

Relationale Begriffsbildungen

Die in Kapitel 1 besprochenen Begriffe bilden die Grundlage einer *statistischen Betrachtungsweise* von Gesamtheiten: Man geht von Eigenschaften aus, die sich den Elementen einer Gesamtheit jeweils individuell zurechnen lassen, und betrachtet dann deren Häufigkeitsverteilungen. In einem gewissen Spannungsverhältnis dazu stehen *relationale Betrachtungsweisen*, die von Beziehungen zwischen den Elementen einer Gesamtheit ausgehen. Damit beschäftigen wir uns in diesem Kapitel. Zuerst werden einige Grundbegriffe besprochen, dann folgen Überlegungen zu relationalen Strukturbegriffen und zu einigen Varianten personeller Netzwerke.

2.1 Relationale Variablen

1. Relationale Aussagen. Von Beziehungen bzw. Relationen wird in unterschiedlichen Bedeutungen geredet, einige Unterscheidungen werden im Abschnitt 2.2 besprochen. Hier soll zunächst angenommen werden, dass man ohne weiteres relationale Aussagen formulieren kann, zum Beispiel: Zwei Menschen kennen sich oder sind befreundet oder sind verheiratet;

ein Mensch erzielt ein höheres Einkommen als ein anderer; zwei Schüler sind Mitglieder derselben Schulklasse; ein Mensch ist Angestellter eines bestimmten Unternehmens; ein Unternehmen bezieht von einem anderen Unternehmen Vorleistungen für seine Güterproduktion; zwei Computer sind durch ein Netzwerk verbunden, so dass Daten ausgetauscht werden können. Dies sind Beispiele für *relationale Aussagen*: Aussagen, die sich gleichzeitig auf zwei (oder mehr) Objekte beziehen. Zu unterscheiden sind relationale Aussagen und Aussageformen. Zum Beispiel ist 'Franz ist verheiratet mit Karin' eine *relationale Aussage*, die ihrer Intention nach einen Sachverhalt ausdrückt und infolgedessen wahr oder falsch sein kann. Dagegen ist 'ω ist verheiratet mit ω'' eine *relationale Aussageform*, wenn ω und ω' nicht Namen bestimmter Objekte, sondern *logische Variablen* sind. Relationale Aussagen, die wahr oder falsch sein können, entstehen erst dann, wenn man in die logischen Variablen (Leerstellen) bestimmte Namen einsetzt (z.B. Franz und Karin).

Im Folgenden soll das Symbol $\sim$ dazu dienen, um auf *relationale Ausdrücke* zu verweisen. Wenn man inhaltlich bestimmte Aussagen machen möchte, muss natürlich eine Bedeutung vereinbart werden. Zum Beispiel könnte vereinbart werden, dass das Symbol $\sim$ bis auf weiteres als Abkürzung für den relationalen Ausdruck 'ist verheiratet mit' verwendet werden soll. Unabhängig von der Vereinbarung einer bestimmten Bedeutung können jedoch mit dem Symbol $\sim$ relationale Aussageformen formuliert werden, die allgemein die Form $\omega \sim \omega'$ haben. In dieser Schreibweise handelt es sich also um eine Aussageform. Erst wenn man dem Symbol $\sim$ eine bestimmte Bedeutung gibt und anstelle von ω und ω' Namen für bestimmte Objekte einsetzt, entsteht eine relationale Aussage, die wahr oder falsch sein kann.

2. Ein expliziter Relationsbegriff. Offenbar muss überlegt werden, auf welche Arten von Objekten man sich beziehen kann, um aus relationalen Aussageformen relationale Aussagen zu machen. Die Umgangssprache orientiert sich an der Bedeutung der relationalen Ausdrücke. Ist zum Beispiel für das Symbol $\sim$ die Bedeutung 'ist verheiratet mit' vereinbart worden, ist klar, dass man nur dann zu sinnvollen Aussagen gelangt, wenn man für ω und ω' Namen von Menschen einsetzt. Für die weiteren Überlegungen soll angenommen werden, dass man sich jeweils auf eine explizit definierte Menge beziehen kann, deren Elemente als Objekte für relationale Aussagen verwendet werden können. Zur symbolischen Repräsentation dient wie bei der Definition statistischer Variablen die Schreibweise $\Omega := \{\omega_1, \ldots, \omega_n\}$. Wie in Abschnitt 1.1 erläutert wurde, sind $\omega_1, \ldots, \omega_n$ (fiktive) Namen für die Objekte, auf die man sich gedanklich beziehen möchte, und das Symbol Ω dient zum Verweis auf die Menge dieser Namen bzw. Objekte.

Nach diesen Vorüberlegungen kann der Begriff einer Relation, wie er im weiteren verwendet werden soll, explizit definiert werden. Eine *Relation* besteht aus drei Bestandteilen:

- Es muss ein relationaler Ausdruck $\sim$ eingeführt werden, mit dem relationale Aussageformen der Gestalt $\omega \sim \omega'$ gebildet werden können. (Sobald man nicht nur rein formale Betrachtungen anstellen möchte, muss natürlich auch die inhaltliche Bedeutung angegeben werden.)
- Es muss eine Objektmenge $\Omega := \{\omega_1, \ldots, \omega_n\}$ angegeben werden, deren Elemente als Namen verwendet werden können, um relationale Aussagen zu bilden.
- Schließlich muss angegeben werden, welche der insgesamt möglichen relationalen Aussagen wahr bzw. falsch sind.

Es wäre also eine verkürzte und potentiell irreführende Redeweise, das Symbol $\sim$ eine Relation zu nennen. Dieses Symbol bildet nur ein Hilfsmittel zur Formulierung relationaler Aussagen. Die Relation selbst besteht vielmehr in der Gesamtheit der zutreffenden relationalen Aussagen, die man mithilfe des relationalen Ausdrucks $\sim$ über alle möglichen Paare von Objekten in der Objektmenge Ω machen kann. Sobald man sich dies klargemacht hat, kann man natürlich von einer Relation $(\Omega, \sim)$ sprechen und auch abkürzend von einer Relation $\sim$, wenn der Bezug auf eine bestimmte Objektmenge durch den Kontext gegeben ist.

Ein einfaches Beispiel kann die Begriffsbildungen illustrieren. Die Objektmenge besteht aus 5 Personen: $\Omega := \{\omega_1, \ldots, \omega_5\}$, und es soll festgestellt werden, wer mit wem verheiratet ist. Die Bedeutung des Symbols $\sim$ wird also durch 'ist verheiratet mit' festgelegt. Mithilfe der Aussageform $\omega \sim \omega'$ können in diesem Beispiel auf insgesamt 25 unterschiedliche Weisen relationale Aussagen gebildet werden. Einige davon sind richtig, die übrigen sind falsch. Angenommen, dass ω_1 und ω_3 und ω_2 und ω_4 verheiratet sind, gibt es folgende Aussagen:

Zutreffende Aussagen	Unzutreffende Aussagen		
$\omega_1 \sim \omega_3$	$\omega_1 \sim \omega_1$	$\omega_2 \sim \omega_5$	$\omega_4 \sim \omega_4$
$\omega_3 \sim \omega_1$	$\omega_1 \sim \omega_2$	$\omega_3 \sim \omega_2$	$\omega_4 \sim \omega_5$
$\omega_2 \sim \omega_4$	$\omega_1 \sim \omega_4$	$\omega_3 \sim \omega_3$	$\omega_5 \sim \omega_1$
$\omega_4 \sim \omega_2$	$\omega_1 \sim \omega_5$	$\omega_3 \sim \omega_4$	$\omega_5 \sim \omega_2$
	$\omega_2 \sim \omega_1$	$\omega_3 \sim \omega_5$	$\omega_5 \sim \omega_3$
	$\omega_2 \sim \omega_2$	$\omega_4 \sim \omega_1$	$\omega_5 \sim \omega_4$
	$\omega_2 \sim \omega_3$	$\omega_4 \sim \omega_3$	$\omega_5 \sim \omega_5$

Die Relation besteht in diesem Beispiel aus der Gesamtheit der 25 Aussagen, von denen 4 zutreffend, die übrigen 21 nicht zutreffend sind.

Das Beispiel zeigt, dass sich eine Relation auf alle möglichen Paare von Objekten bezieht, die man aus den Elementen einer Objektmenge bilden kann. Um diese Paare zu bilden, verwendet man in der Mengenlehre den Begriff eines kartesischen Produkts. Bezieht man sich allgemein auf zwei Mengen A und B, besteht ihr *kartesisches Produkt*, geschrieben $A \times B$, aus allen geordneten Paaren der Form (a, b), wobei a ein Element aus A und b ein Element aus B ist. Ist zum Beispiel $A = \{a_1, a_2, a_3\}$ und $B = \{b_1, b_2\}$, erhält man: $A \times B = \{(a_1, b_1), (a_1, b_2), (a_2, b_1), (a_2, b_2), (a_3, b_1), (a_3, b_2)\}$. Bei endlichen Mengen gilt offenbar $|A \times B| = |A| \cdot |B|$, wobei, wenn M irgendeine endliche Menge ist, mit dem Ausdruck $|M|$ auf die Anzahl ihrer Elemente verwiesen werden soll.

Eine Relation für eine Objektmenge Ω gibt nun offenbar für jedes Element $(\omega, \omega') \in \Omega \times \Omega$ an, ob die relationale Aussage $\omega \sim \omega'$ zutrifft oder nicht. Somit kann man auch jede Relation für eine Objektmenge Ω durch eine Teilmenge des kartesischen Produkts $\Omega \times \Omega$ festlegen, die genau diejenigen Paare (ω, ω') enthält, für die die relationale Aussage zutrifft. In unserem Beispiel: $R^* := \{(\omega_1, \omega_3), (\omega_3, \omega_1), (\omega_2, \omega_4), (\omega_4, \omega_2)\}$. Diese Methode wird *Definition einer Relation durch ein kartesisches Produkt (einer Objektmenge Ω mit sich selbst)* genannt. Offenbar entspricht jeder Teilmenge von $\Omega \times \Omega$ eine spezifische Relation für die Elemente von Ω.

3. Relationale Variablen. Eine andere Möglichkeit, um sich begrifflich auf Relationen für eine Objektmenge Ω zu beziehen, besteht in der Verwendung *relationaler Variablen*. Mit diesem Begriff sind zunächst (später wird die Definition verallgemeinert) Funktionen gemeint, die folgende Form haben: $R : \Omega \times \Omega \longrightarrow \{0, 1\}$. R ist der Name der Funktion (der relationalen Variablen), $\Omega \times \Omega$ ist ihr Definitionsbereich, und $\{0, 1\}$ ist ihr Wertebereich. Die Funktion (relationale Variable) R ordnet also jedem Element $(\omega, \omega') \in \Omega \times \Omega$ einen Wert $R(\omega, \omega') \in \{0, 1\}$ zu, wobei folgende Bedeutung vereinbart wird:

$$R(\omega, \omega') := \begin{cases} 1 & \text{wenn } \omega \sim \omega' \text{ zutrifft} \\ 0 & \text{wenn } \omega \sim \omega' \text{ nicht zutrifft} \end{cases}$$

Wie sich später zeigen wird, ist der Begriff einer relationalen Variablen sehr nützlich, weil er sich leicht verallgemeinern lässt, um in komplexerer Weise von Relationen zu sprechen. Außerdem gibt es eine gedanklich einfache Parallele zu statistischen Variablen, also zu Funktionen $X : \Omega \longrightarrow \tilde{\mathcal{X}}$, die jedem Element einer Objektmenge Ω einen Merkmalswert in einem Merkmalsraum $\tilde{\mathcal{X}}$ zuordnen. Der Unterschied besteht nur darin, dass eine statistische Variable jedem einzelnen Objekt, eine relationale Variable dagegen jedem Paar von Objekten einen Merkmalswert zuordnet.

An dieser Parallele knüpft auch eine weitere Möglichkeit zur Darstellung von Relationen an. Beziehen wir uns zunächst auf eine statistische Variable $X : \Omega \longrightarrow \tilde{\mathcal{X}}$. Ihre Werte (die Daten) können in Form einer Datenmatrix dargestellt werden, die folgende Form hat:

ω_1	$X(\omega_1)$
$\vdots$	$\vdots$
ω_n	$X(\omega_n)$

Jede Zeile bezieht sich auf jeweils ein Objekt der Objektmenge Ω. Die erste Spalte enthält den Namen des Objekts, die zweite Spalte den Merkmalswert, der dem Objekt durch die Variable zugeordnet wird. Auf ähnliche Weise kann man die Werte einer relationalen Variablen durch ein zweidimensionales Schema darstellen, das allgemein folgende Form hat:

	ω_1	$\cdots$	ω_n
ω_1	$R(\omega_1, \omega_1)$	$\cdots$	$R(\omega_1, \omega_n)$
$\vdots$	$\vdots$		$\vdots$
ω_n	$R(\omega_n, \omega_1)$	$\cdots$	$R(\omega_n, \omega_n)$

Für das oben angeführte Beispiel erhält man folgende Darstellung:

	ω_1	ω_2	ω_3	ω_4	ω_5
ω_1	0	0	1	0	0
ω_2	0	0	0	1	0
ω_3	1	0	0	0	0
ω_4	0	1	0	0	0
ω_5	0	0	0	0	0

Wenn ein Schema dieser Art verwendet wird, um eine Relation darzustellen, spricht man von einer *Adjazenzmatrix*.

4. Allgemeine relationale Variablen. Als einheitlicher begrifflicher Rahmen zur Erfassung von Beziehungen zwischen den Elementen einer Objektmenge Ω eignen sich relationale Variablen, die in folgender Form als Funktionen definiert sind: $R : \Omega \times \Omega \longrightarrow \tilde{\mathcal{R}}$. Hierbei ist $\tilde{\mathcal{R}}$ ein im Prinzip beliebig konzipierbarer Merkmalsraum. Die relationale Variable R ordnet jedem Element $(\omega, \omega') \in \Omega \times \Omega$ einen Merkmalswert $R(\omega, \omega') \in \tilde{\mathcal{R}}$ zu. Als Spezialfall, wenn $\tilde{\mathcal{R}} = \{0, 1\}$ ist, erhält man einfache relationale Variablen, durch die nur erfasst wird, ob eine Beziehung zwischen zwei Elementen aus Ω besteht. Durch eine Verwendung beliebiger Merkmalsräume können darüber hinaus Beziehungen unterschieden oder näher charakterisiert werden.

Weitere Verallgemeinerungen können sich anschließen. Analog zu mehrdimensionalen statistischen Variablen können mehrdimensionale relationale Variablen $(R_1, \ldots, R_m) : \Omega \times \Omega \longrightarrow \tilde{\mathcal{R}}_1 \times \ldots \times \tilde{\mathcal{R}}_m$ betrachtet werden, durch die für eine Objektmenge Ω gleichzeitig m relationale Variablen $R_1, \ldots, R_m$ mit den zugehörigen Merkmalsräumen $\tilde{\mathcal{R}}_1, \ldots, \tilde{\mathcal{R}}_m$ definiert werden. Außerdem können multi-modale relationale Variablen verwendet werden, um Beziehungen zwischen den Elementen unterschiedlicher Arten von Objektmengen zu erfassen.

Schließlich lässt sich auch eine allgemeine Redeweise von Netzwerken anschließen. Wir verwenden in diesem Text folgende Definition: Etwas ist ein *Netzwerk,* wenn bzw. insoweit es durch eine (ggf. mehrdimensionale und multi-modale) relationale Variable repräsentiert werden kann. – Das unspezifische Reden von „etwas“ soll es erlauben, sich mit dem Netzwerkbegriff nicht nur auf materielle Aspekte unserer Erfahrungswelt zu beziehen, sondern auch auf abstrakte Mengen wie z.B. Zahlen oder Merkmalsräume. Für diesen allgemeinen Netzwerkbegriff werden somit keine besonderen Anforderungen an die Beschaffenheit der Objekte (die auch als *Knoten* des Netzwerks bezeichnet werden) oder der sie verknüpfenden Beziehungen gestellt. Insofern für das Reden von Netzwerken eine bestimmte formale Repräsentation vorausgesetzt wird, ist der Begriff dennoch enger als der abstrakte Systembegriff.

2.2 Relationale Strukturbegriffe

Der Strukturbegriff wird hauptsächlich in zwei unterschiedlichen Bedeutungen verwendet: Einerseits in einer statistischen Bedeutung, in der sich das Wort auf eine oder mehrere Merkmalsverteilungen in einer statistischen Gesamtheit bezieht; damit haben wir uns in Abschnitt 1.2 beschäftigt. Andererseits wird das Wort verwendet, um in einer vergleichsweise unspezifischen Weise auf die Gliederung und den Aufbau irgendeines (realen oder fiktiven) Sachverhalts zu verweisen. In dieser zweiten Bedeutung sagte zum Beispiel Wittgenstein in seinem *Tractatus* (2.032): „Die Art und Weise, wie die Gegenstände im Sachverhalt zusammenhängen, ist die Struktur des Sachverhalts.“ In einer ähnlichen Formulierung heißt es bei G. Hernes (1976: 518): „A structure is a configuration of parts, and a structural description is a characterization of the way the components in a set are interrelated.“ An dieses Wortverständnis knüpfen die relationalen (Sozial-) Strukturbegriffe an, mit denen wir uns in diesem Abschnitt beschäftigen.

1. Definition eines relationalen Strukturbegriffs. Zu einer allgemeinen Definition eines relationalen Strukturbegriffs gelangt man, wenn man sich auf Gesamtheiten von Objekten bezieht, die durch Beziehungen miteinander verbunden sind. In diesem Kontext bezieht sich das Wort darauf, wie die Objekte durch Beziehungen zusammenhängen, und wir sprechen dann von einem *relationalen Strukturbegriff.*

In dieser allgemeinen Bedeutung kann der relationale Strukturbegriff auf beliebige Systeme (im Sinne des abstrakten Systembegriffs) angewendet werden, insbesondere auf Netzwerke, so wie dieser Begriff in Abschnitt 2.1 (§ 4) definiert worden ist. Allerdings muss auf eine Ambivalenz in der Begriffsverwendung geachtet werden. Wenn man den relationalen Strukturbegriff in seiner allgemeinen Bedeutung verwendet, gibt es zunächst keinen begrifflichen Unterschied zwischen dem Netzwerk und seiner Struktur; oder anders formuliert: Man beschreibt die Struktur eines Netzwerks, indem man das Netzwerk beschreibt.[1] Um diese Ambivalenz zu vermeiden, werden wir – in Übereinstimmung auch mit den anfangs angeführten Zitaten von Wittgenstein und Hernes – festlegen, dass sich der relationale

[1]So heißt es z.B. bei D. Krackhardt (1987: 113): „The structure of any system is defined as a set of relational statements between all pairs of actors in the system.“

Strukturbegriff nur auf formale Eigenschaften eines Netzwerks (oder Systems) bezieht, beispielsweise auf die Anzahl der Objekte und den Grad der Dichte ihrer Beziehungen, um nur zwei formale Aspekte zu nennen, nicht jedoch auf die Art der Objekte und die inhaltliche Bedeutung der Beziehungen.[2] Infolgedessen ist es auch möglich, dass unterschiedliche Netzwerke die gleiche relationale Struktur aufweisen.

2. Unterschiedliche Arten von Beziehungen. Um Netzwerke zu verstehen, ist nicht nur ihre Struktur von Bedeutung; zuerst muss man die Knoten und Beziehungen verstehen, auf die das Netzwerk Bezug nimmt. Darin unterscheiden sich auch zunächst die Netzwerke, die in der empirischen Sozialforschung betrachtet werden können. Allerdings ist es kaum möglich, eine vollständige Übersicht über alle Möglichkeiten zu gewinnen, in denen von Beziehungen gesprochen werden kann. Wir beschränken uns deshalb an dieser Stelle darauf, einige gelegentlich nützliche allgemeine Unterscheidungen anzudeuten.[3]

a) Eine Möglichkeit, von Beziehungen zwischen zwei oder mehr Objekten zu sprechen, beruht auf einem Vergleich von Eigenschaften, durch die man die Objekte zunächst jeweils separat charakterisieren kann; wir sprechen dann von *komparativen Beziehungen*. Zum Beispiel: ω ist größer als ω' oder ist älter als ω' oder ist gleichalt wie ω'. Insbesondere kann man solche Beziehungen bilden, wenn zunächst eine statistische Variable gegeben ist, indem man die Merkmalswerte von zwei oder mehr Objekten vergleicht.

b) Eine andere Möglichkeit zur Definition von Beziehungen zwischen zwei oder mehr Objekten besteht darin, auf eine Situation oder einen Kontext Bezug zu nehmen, dem die Objekte in einer bestimmten Weise angehören. So gelangt man zu *kontextabhängigen Beziehungen*. Zur Charakterisierung solcher Beziehungen kann man sowohl von Ereignissen als auch von Sachverhalten ausgehen.

   (α) Einerseits kann man von *ereignisförmigen Beziehungen* zwischen zwei oder mehr Objekten sprechen, wenn die Objekte irgendwie in ein die Beziehung konstituierendes Ereignis einbezogen sind. Zum Beispiel zwei Personen, die sich unterhalten, bei einem Verkehrsunfall zusam-

[2]Dieser Vorschlag zum Sprachgebrauch entspricht auch folgender Bemerkung von F. U. Pappi (1987: 15): „Netzwerke sind nach unserer Definition empirische Systeme. Sie lassen sich formal als Graphen darstellen.“ Etwas ausführlicher heißt es bei K.-D. Opp und H. J. Hummell (1973: 67): Es „soll im folgenden unter *Struktur* eine spezielle formale Eigenschaft von Netzwerken (Relationengebilden) verstanden werden. Beschreibt man zwei Relationengebilde durch Boolesche Matrizen [= Adjazenzmatrizen], dann sollen die *Netzwerke von gleicher Struktur* genau dann sein, wenn die zugeordneten Matrizen identisch bzw. durch Permutationsmatrizen ineinander transformierbar sind.“

[3]In der Literatur, die sich mit sozialen Netzwerken beschäftigt, beziehen sich zahlreiche Überlegungen auch auf inhaltliche Unterscheidungen; man vgl. die Diskussion bei F. U. Pappi (1987: 16ff.).

menstoßen oder an der gleichen Landtagswahl teilgenommen haben. Das zuletzt angeführte Beispiel zeigt, dass eine ereignisförmige Beziehung zwischen zwei Objekten nicht impliziert, dass es zwischen den Objekten auch einen direkten (oder allgemeiner: irgendwie kausal relevanten) Kontakt gibt. Wir sprechen deshalb im engeren Sinn von einer *Interaktionsbeziehung* (oder kurz: *Interaktion*), wenn in irgendeiner Form ein kommunikativer Austausch und/oder eine physische Wechselwirkung stattfindet.[4] Natürlich kann man zur Definition ereignisförmiger Beziehungen auch zeitliche Sequenzen mehrerer Ereignisse verwenden; man denke z.B. an Beschreibungen persönlicher Beziehungen, bei denen fast immer solche Bezugnahmen auf ihre Geschichte stattfinden. In jedem Fall, auch wenn nur auf ein Ereignis Bezug genommen wird, setzen empirische Feststellungen über ereignisförmige Beziehungen eine retrospektive Betrachtungsweise voraus.

(β) Andererseits kann man zur Definition von Beziehungen auch von Sachverhalten ausgehen, bei denen es sich nicht um Ereignisse handelt; zum Beispiel: Zwei Orte sind durch eine Straße miteinander verbunden. Es ist allerdings fraglich, ob der Unterscheidung zwischen Ereignissen und Sachverhalten auch eine relevante Unterscheidung zwischen Arten von Beziehungen entspricht. Denn in einer allgemeinen Sprechweise kann man auch dann von einem Sachverhalt sprechen, wenn ein Ereignis stattgefunden hat; und andererseits impliziert die gedankliche Bezugnahme auf einen Sachverhalt nicht, dass er während eines längeren Zeitraums (unverändert) existiert. Das mag der Fall sein (wie vermutlich bei der Straße zwischen den beiden Orten), aber es muss nicht der Fall sein (wie z.B. bei zwei Computern, die nur für einen kurzen Zeitraum durch ein Kabel miteinander verbunden werden, um Daten zu übertragen).

Anhand von Beispielen kann man sich verdeutlichen, dass es auch zwischen komparativen und kontextabhängigen Beziehungen keine vollständig scharfe Unterscheidung gibt. Ein wichtiges Beipiel ist räumliche Nähe, die man sowohl als eine komparative als auch als eine kontextabhängige Beziehung betrachten kann.

3. Faktische und modale Betrachtungsweisen. Wichtiger als allgemeine Unterscheidungen zwischen Arten von Beziehungen ist der Umstand, dass es in vielen Fällen bei ihrer Betrachtung und Darstellung eine wesentliche Ambivalenz gibt. Zum Beispiel: Zwei Computer sind durch ein Kabel für den Austausch von Daten miteinander verbunden. Das ist einerseits eine empirisch feststellbare Tatsache, die jedoch andererseits auf eine Möglich-

[4] Der in diesem Text verwendete Begriff einer Interaktion setzt also nicht unbedingt Akteure voraus; und auch dann, wenn er sich auf individuelle Akteure bezieht, muss es sich nicht um „soziales Handeln“ im Sinne Max Webers oder um „bewertende Austauschprozesse“ wie etwa bei George Homans (1961: 35) handeln.

keit, nämlich einen Austausch von Daten, verweist. Oder: Eine Person ist bei einem Unternehmen angestellt. Diese Feststellung verweist einerseits auf eine bestimmte Tatsache – dass irgendwann ein Arbeitsvertrag vereinbart wurde und immer noch besteht –, andererseits auf mögliche Verhaltensweisen, die infolge des Arbeitsvertrags realisiert werden sollten. Wie in diesen Beispielen kann man in vielen Fällen zwei Aspekte unterscheiden:

a) Einerseits einen *faktischen* Aspekt, der sich auf empirisch fixierbare Sachverhalte oder Ereignisse (z.B. eine Kabelverbindung oder das Vorhandensein eines Arbeitsvertrags) bezieht, durch die eine Beziehung faktisch begründet wird; und

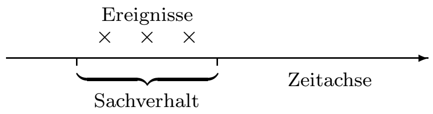
b) andererseits einen *modalen* Aspekt, der sich – je nach der Art der beteiligten Objekte oder Personen – auf mögliche Verhaltensweisen bezieht, die infolge der faktischen Beziehung möglich oder wahrscheinlich oder normativ gefordert werden.

Es handelt sich dabei nicht um unterschiedliche Arten von Beziehungen, sondern um unterschiedliche Betrachtungsweisen einer Beziehung. Um noch ein Beispiel anzuführen: Die Aussage, dass zwei Personen verheiratet sind, kann einerseits bedeuten, dass zu einem bestimmten Zeitpunkt ein Ereignis stattgefunden hat, durch das die beiden Personen verheiratet wurden. Andererseits können aber auch bestimmte Verhaltensweisen gemeint sein, die durch dieses Ereignis möglich und/oder normativ gefordert werden.

Zu betonen ist, dass sich die modale Betrachtungsweise auf Möglichkeiten bezieht. Zwar kann man bei allen Beziehungen, die eine modale Betrachtungsweise erlauben, auch eine *retrospektive* Betrachtungsweise einnehmen (deren Reichweite natürlich von der bisherigen Dauer der Beziehung abhängt). So kann man z.B. feststellen, in welchem Umfang Daten zwischen den beiden Computern ausgetauscht worden sind oder wie sich die beiden Personen während ihrer bisherigen Ehe zueinander verhalten haben. Bei der modalen Betrachtungsweise geht es jedoch nicht um eine retrospektive Feststellung von Interaktionen, sondern darum, *wie durch Beziehungen mögliche Verhaltensweisen der jeweils beteiligten Objekte konstituiert werden.* Dabei zielt das Erkenntnisinteresse nicht nur auf Unterschiede in den Verhaltensweisen selbst, sondern auch darauf, wie sie durch die Beziehung ermöglicht, wahrscheinlich gemacht oder normativ gefordert werden.

4. Beziehungen und mögliche Ereignisse. Denken wir noch einmal an die beiden Computer. Offenbar kann der Sachverhalt, dass sie durch ein Kabel miteinander verbunden sind, verwendet werden, um eine Beziehung zwischen ihnen festzustellen. Natürlich könnten auch zahlreiche andere Sachverhalte verwendet werden, um andere Beziehungen zwischen ihnen festzustellen; zum Beispiel, dass sie nebeneinander auf einem Tisch stehen oder sich im gleichen Raum befinden. Bei den Beziehungen, die auf diese Weise definiert werden können, handelt es sich ersichtlich nicht um ereignisförmige Beziehungen, insbesondere nicht um Interaktionen.

Ein Zusammenhang kann jedoch hergestellt werden, wenn die durch einen Sachverhalt definierte Beziehung eine modale Betrachtungsweise erlaubt. In unserem Beispiel ist das der Fall, wenn die beiden Computer durch ein Kabel verbunden sind, so dass es möglich wird, Daten auszutauschen. Wie das folgende Bild andeutet, ermöglicht dann der die Beziehung definierende Sachverhalt Ereignisse einer bestimmten Art:



Die Ereignisse (Austausch von Daten) werden durch den Sachverhalt (die Kabelverbindung) nicht verursacht, sondern ermöglicht. Insofern bilden sie keinen realen, sondern nur einen modalen Aspekt der durch den Sachverhalt definierten Beziehung zwischen den beiden Computern. Wenn jedoch solche Ereignisse stattfinden, kann man sich retrospektiv auf sie beziehen und dadurch auch eine ereignisförmige Beziehung definieren. Sie ist natürlich mit der ursprünglich durch den Sachverhalt definierten Beziehung nicht identisch.

Formal analog verhält es sich, wenn ereignisförmige Beziehungen eine modale Betrachtungsweise erlauben, also als Bedingungen für mögliche spätere Ereignisse betrachtet werden können. Das Basisereignis, durch das die Beziehung zunächst zustande kommt, kann dann nämlich als zeitlicher Beginn eines Sachverhalts aufgefasst werden, während dessen Vorhandensein wie im oben skizzierten Bild weitere Ereignisse stattfinden können. Zum Beispiel besteht das Basisereignis darin, dass ein Arbeitsvertrag abgeschlossen wird; und dadurch entsteht dann für eine gewisse Zeit ein bestimmter Sachverhalt, der seinerseits einen Rahmen für weitere Ereignisse (anderer Art) bildet.

5. Soziale Beziehungen. In der sozialwissenschaftlichen Literatur wird in unterschiedlichen Bedeutungen von „sozialen Beziehungen" gesprochen. Soziologen orientieren sich oft an folgender Definition Max Webers:

> „Soziale „Beziehung" soll ein seinem Sinngehalt nach aufeinander gegenseitig *eingestelltes* und dadurch orientiertes Sichverhalten mehrerer heißen. Die soziale Beziehung *besteht* also durchaus und ganz ausschließlich: in der *Chance*, daß in einer (sinnhaft) angebbaren Art sozial gehandelt wird, einerlei zunächst: worauf diese Chance beruht." (Weber 1921/1976: 13)

Diese Definition sozialer Beziehungen ist jedoch aus mehreren Gründen problematisch.

– Zunächst deshalb, weil sie von vornherein nur personelle Beziehungen einbezieht und, noch enger, nur „soziales Handeln" in Betracht zieht. Somit werden viele wichtige Arten von Beziehungen, wie z.B. Beziehungen zwischen Organisationen oder indirekte Abhängigkeitsbeziehungen zwischen Personen, die nicht durch „sinnhafte Orientierungen" erschlossen werden können, nicht erfasst.

– Ein zweites Problem betrifft das Verständnis faktischer und modaler Aspekte von Beziehungen. Webers Definition bezieht sich zunächst auf einen faktischen Aspekt („Sichverhalten"), wechselt dann aber durch ein Reden von „Chancen" unvermittelt in eine scheinbar modale Betrachtungsweise.

– Schließlich liefert aber diese modale Betrachtungsweise gerade bei Beziehungen zwischen sozialen Akteuren in den meisten Fällen kein adäquates Verständnis. Denn Webers Chancenbegriff bezieht sich auf „statistische Wahrscheinlichkeiten" und infolgedessen auf Häufigkeiten von Verhaltensweisen, die bei einer retrospektiven Betrachtung von Beziehungen festgestellt werden können.[5] Die wesentliche Differenz zwischen einer modalen und einer retrospektiven Betrachtungsweise von Beziehungen wird infolgedessen nicht nur verwischt, sondern verkehrt. Folgte man Webers Definition, ergäbe sich z.B. der Inhalt der Beziehung, die zwischen einer Person und einem Unternehmen durch den Abschluss eines Arbeitsvertrags zustande kommt, durch das nachfolgende Verhalten der Person und des Unternehmens (denn auf dieses Verhalten bezieht sich Webers Chancenbegriff). Eine modale Betrachtungsweise dieser Beziehung müsste dagegen auf die normativen Festlegungen des Arbeitsvertrags Bezug nehmen.

Besonders wichtig erscheint mir der zuletzt genannte Kritikpunkt. Denn dabei geht es nicht nur um ein richtiges Verständnis modaler Betrachtungsweisen von Beziehungen (insbesondere zwischen sozialen Akteuren) und ihre Unterscheidung von retrospektiven Betrachtungsweisen.[6] Vielmehr geht es auch um die Frage, ob und ggf. wie man zumindest in einigen Fällen Beziehungen als Bedingungen für das Verhalten beteiligter Akteure betrachten kann. Webers Ansatz verstellt schon den Zugang zu dieser Fragestellung, denn durch empirische Häufigkeiten definierte „Chancen" für das je faktische Verhalten (im Kontext einer Beziehung) können nicht als dessen Bedingungen aufgefasst werden.

Es ist jedoch nicht erforderlich, Webers Definition durch eine andere

---

[5] Insbesondere bezieht sich Webers Chancenbegriff nicht auf Handlungschancen, die von Akteuren reflektiert und wahrgenommen werden können. Zur Unterscheidung zwischen Handlungschancen und „statistischen Chancen" vgl. man die Ausführungen bei Rohwer und Pötter (2002b: 166ff.).

[6] Man kann sich das Problem an beliebig vielen Beispielen verdeutlichen. Angenommen, ω hat mit ω' einen Kaufvertrag abgeschlossen, die vereinbarte Ware aber nicht geliefert. Sollte man dann (unter Berufung auf Webers Definition) sagen, dass ω' die durch den Kaufvertrag begründete Beziehung zu ω missverstanden habe, da offenbar das „wahrscheinliche" Verhalten von ω falsch eingeschätzt wurde?

zu ersetzen, denn es ist möglich, ganz ohne einen speziellen Begriff „sozialer“ Beziehungen auszukommen. Wichtig ist vielmehr, jeweils deutlich zu machen, welche Arten von Beziehungen betrachtet werden sollen und wie das geschehen soll, nämlich in einer faktischen (retrospektiven) oder modalen Betrachtungsweise.

6. Soziale Netzwerke und relationale Sozialstrukturbegriffe. Mit dem relationalen Strukturbegriff kann man sich auf beliebige Mengen von Objekten beziehen, zwischen denen es irgendwelche Beziehungen gibt, insbesondere auf beliebige Netzwerke. Einschränkend sprechen wir von *sozialen Netzwerken*, wenn es sich bei den Objekten des Netzwerks um *soziale Akteure* (Personen und/oder Organisationen, insbesondere Haushalte und Unternehmen) handelt.[7]

Diese Definition impliziert, dass sich die Sozialforschung nicht nur mit sozialen Netzwerken beschäftigt. Denn erstens sind nicht alle für Fragen der Sozialforschung relevanten Netzwerke im Sinne der obigen Definition soziale Netzwerke, man denke z.B. an Verkehrsverbindungen zwischen Städten. Außerdem gibt es zahlreiche Aspekte gesellschaftlicher Verhältnisse, die nicht (im Sinne unserer formalen Definition) als Netzwerke beschrieben werden können, z.B. statistische Strukturen und Institutionen. Deshalb werden wir auch vermeiden, irgendeinen allgemeinen relationalen Sozialstrukturbegriff zu definieren, wie dies von einigen Autoren versucht worden ist.[8] Stattdessen beziehen wir relationale Strukturbegriffe immer nur auf diejenigen Aspekte gesellschaftlicher Verhältnisse, die durch Netzwerke explizit repräsentiert werden können.[9]

[7]Dies entspricht folgender Definition von S. Wasserman und K. Faust (1994: 20): „A *social network* consists of a finite set or sets of actors and the relation or relations defined on them.“

[8]In der Literatur wird oft auf einen Vortrag Alfred R. Radcliffe-Browns aus dem Jahr 1940 Bezug genommen: „For a preliminary definition of social phenomena it seems sufficiently clear that what we have to deal with are relations of association between individual organisms. [...] Let us consider what are the concrete, observable facts with which the social anthropologist is concerned. If we set out to study, for example, the aboriginal inhabitants of a part of Australia, we find a certain number of individual human beings in a certain natural environment. We can observe acts of behaviour of these individuals, including, of course, their acts of speech, and the material products of past actions. We do not observe a ‘culture’, since that word denotes, not any concrete reality, but an abstraction, and as it is commonly used a vague abstraction. But direct observation does reveal to us that these human beings are connected by a complex network of social relations. I use the term ‘social structure’ to denote this network of actually existing relations. It is this that I regard it as my business to study if I am working, not as an ethnologist or psychologist, but as a social anthropologist.“ (Radcliffe-Brown 1940: 189f.)

[9]Natürlich spricht nichts dagegen, diese Aspekte auch als „Aspekte einer Sozialstruktur“ zu bezeichnen; so kann man z.B. folgende Formulierung von F. U. Pappi (1987: 12) verstehen: „Für den Soziologen ist die Netzwerkanalyse eine Methode zur Untersuchung von sozialen Strukturen. Eine Sozialstruktur wird repräsentiert durch die Beziehungen zwischen sozialen Einheiten wie Personen, Gruppen, Organisationen usw.“

7. Sind relationale Strukturen zeitlich stabil? Einige der Fragen, die in Abschnitt 1.2 im Hinblick auf statistische Strukturen besprochen wurden, stellen sich gleichermaßen für relationale Strukturen. Das betrifft zunächst die Frage der zeitlichen Stabilität. Die in Abschnitt 1.2 (§ 2) zitierten Annahmen über zeitliche Stabilität beziehen sich fast immer auch auf relationale Strukturen. Dementsprechend definierte ein früher Vertreter der Netzwerkanalyse, Edward O. Laumann (1966: 3), „Sozialstruktur“ als ein „persistent system of social relationships among social positions“.[10]

Wie für statistische gilt jedoch auch für relationale Strukturen: dass ihr Begriff keinerlei Annahmen über ihre zeitliche Stabilität impliziert. Somit kann man auch stets fragen, wie sich relationale Strukturen im Zeitablauf verändern. In einer retrospektiven Betrachtung mag sich dann zeigen, dass sich einige Strukturen schneller, andere langsamer verändert haben; aber auch abgesehen davon, dass dies bestenfalls im Nachhinein festgestellt werden kann,[11] ergeben sich daraus keine Einschränkungen für ein Reden von Strukturen.

8. Wie entstehen relationale Strukturen? Auch die Frage, wie Strukturen entstehen, stellt sich für statistische und relationale Strukturen in analoger Weise. So wie in Abschnitt 1.2 bei statistischen Strukturen können auch bei relationalen Strukturen drei Aspekte unterschieden werden:

– Man kann zunächst an die *substantiellen Prozesse* denken, durch die in der sozialen Realität die jeweils thematisierten Beziehungen entstanden sind oder, in einer modalen Betrachtungsweise, entstehen könnten.

– Man kann weiterhin an die *datenerzeugenden Prozesse* denken, durch die Informationen (Daten) über in der sozialen Realität als gegeben vorausgesetzte Beziehungen entstehen.

– Und man kann schließlich an die gedanklichen und rechnerischen Prozesse denken, durch die aus den Daten bestimmte Netzwerke, Charakterisierungen und Modelle konstruiert werden.[12]

Offenbar interessieren in erster Linie die substantiellen Prozesse, und man kann auch sogleich feststellen: Wie diese Prozesse aufzufassen und begrifflich zu konzipieren sind, hängt vor allem von der Art der Beziehungen ab, deren Entstehen überlegt werden soll. Zum Beispiel: Wie entstehen

[10]Diese Definition wurde auch von anderen Autoren übernommen, so etwa von P. V. Marsden und N. Lin (1982: 9).

[11]Deshalb sind Formulierungen der folgenden Art offenbar problematisch: „*Beziehungen* entstehen, sobald Menschen in relativ stabile, kontinuierliche Muster spezifischer Interaktionen und/oder gegenseitiger Abhängigkeit eintreten [...].“ (Joas 2001: 16) Oder: „Soziale Beziehungen sind beständige Interaktionsmuster zwischen zwei oder mehr Personen.“ (Weymann 2001: 104)

[12]Es erscheint durchaus angemessen, hier von einer *Konstruktion* zu sprechen; denn bei der Frage, welche Knoten in die Definition eines Netzwerks einbezogen und welche Beziehungen betrachtet werden sollen, sind mehr oder weniger willkürliche Entscheidungen kaum zu vermeiden.

Verkehrsverbindungen zwischen Städten? Wie entstehen Verkehrsunfälle, durch die zwei oder mehr Menschen in eine physische Interaktionsbeziehung geraten? Wie entstehen Arbeitsverträge, durch die Menschen zu Mitarbeitern eines Unternehmens werden? Wie entstehen Freundschaften? Wie werden zwei Personen zu Mitgliedern derselben Schulklasse oder zu Teilnehmern desselben Seminars?

Die Liste solcher Fragestellungen könnte fast beliebig fortgesetzt werden. Bemerkenswert ist vor allem, dass es keine allgemeine Prozesskonzeption gibt, die sich gleichermaßen für alle Fragestellungen eignet. In einigen Fällen erscheint es sinnvoll, an Handlungsprozesse zu denken, an denen zwei oder mehr Menschen beteiligt sind; aber eine solche Vorstellung passt nicht immer, denn eine Beziehung kann auch dadurch enstehen, dass zwei oder mehr Prozesse zunächst unabhängig voneinander ablaufen, bevor sie irgendwann tatsächlich zu einer Interaktion führen oder auch nur zu einer Situation, die zur Feststellung einer komparativen Beziehung verwendet werden kann. Weiterhin kann man auch an Prozesse denken, die gar nicht als Handlungsprozesse verstanden werden können, wie z.B. die Ausbreitung von Krankheiten durch eine Übertragung von Viren oder Bakterien. Wir kommen also zu dem Ergebnis, dass eine allgemeine Antwort auf die Frage, wie relationale Strukturen entstehen, nicht gegeben werden kann.

2.3 Varianten personeller Netzwerke

1. Personelle und personell konstituierte Netzwerke. Bei der Diskussion personeller Netzwerke ist folgende Unterscheidung sinnvoll: Einerseits gibt es Netzwerke, deren Knoten sich auf individuelle Personen beziehen; dann sprechen wir von *personellen Netzwerken*. Andererseits gibt es Netzwerke, deren Knoten sich nicht auf individuelle Personen beziehen, bei denen jedoch die Beziehungen zwischen den Knoten durch Personen (nicht unbedingt, aber in vielen Fällen auch durch persönliche Beziehungen) zustande kommen. In diesen Fällen sprechen wir von *personell konstituierten Netzwerken*.

Diese Definition personeller Netzwerke setzt nur voraus, dass sich die Knoten auf Personen (individuelle Akteure) beziehen, lässt es aber offen, welcher Art die Beziehungen sind. Insbesondere impliziert die Definition nicht, dass sich die Personen, zwischen denen eine Beziehung besteht, wechselseitig kennen. Somit können zur Konstruktion personeller Netzwerke sowohl komparative als auch kontextabhängige Beziehungen verwendet werden.[13]

Für die Interpretation personeller Netzwerke ist es offenbar wichtig, ob und ggf. wie die jeweils erfassten Beziehungen eine modale Betrachtungsweise erlauben. Wenn im Alltag von „persönlichen Beziehungen" gesprochen wird, findet meistens eine Bezugnahme sowohl auf faktische Aspekte (man kennt sich, lebt zusammen in einem Haushalt, arbeitet in der gleichen Abteilung usw.) als auch auf modale Aspekte (man kann sich ansprechen, um Rat fragen, um etwas bitten usw.) statt.[14]

[13] Es sei angemerkt, dass in der Literatur oft in einem engeren Sinn von „personellen Netzwerken" gesprochen wird, bei dem vorausgesetzt wird, dass es sich um persönliche Beziehungen zwischen den beteiligten Personen handelt.

2. Durch Ereignisse definierte personelle Netzwerke. Zur Konstruktion personeller Netzwerke können auch ereignisförmige Beziehungen verwendet werden, deren modale Betrachtung fragwürdig sein kann. Zur Illustration betrachten wir ein in der Literatur oft diskutiertes Beispiel, in dem ein personelles Netzwerk durch eine Folge von Ereignissen definiert wird. Tabelle 2.3-1 zeigt die Daten so, wie sie zuerst von George C. Homans (1951: 83) publiziert wurden.[15] Die Daten beziehen sich auf die Teilnahme von 18 Frauen an sozialen Ereignissen (z.B. Treffen in einem Club oder bei einem kirchlichen Abendessen). Für jedes von 14 zeitlich aufeinander folgenden Ereignissen wird angegeben, welche der Frauen an ihnen teilgenommen haben (in der Tabelle durch ein Kreuz markiert).

Diese Daten können auf zwei unterschiedliche Weisen zur Definition personeller Netzwerken verwendet werden, wobei angenommen wird, dass man sich auf die Gesamtheit der Frauen $\Omega := \{\omega_1, \ldots, \omega_{18}\}$ beziehen möchte:

a) Man kann eine zeitliche Folge von Netzwerken definieren. Für jedes Ereignis $t = 1, \ldots, 14$ gibt es dann eine relationale Variable

$$R_t : \Omega \times \Omega \longrightarrow \{0, 1\}$$

wobei $R_t(\omega, \omega')$ den Wert 1 bekommt, wenn ω und ω' gemeinsam am t-ten Ereignis teilgenommen haben, und andernfalls den Wert 0. Somit gibt es in diesem Fall eine Folge von 14 Adjazenzmatrizen $\mathbf{A}_t$.

b) Stattdessen kann man auch die Teilnahme an allen Ereignissen betrachten und Beziehungen durch die Anzahl der Ereignisse definieren, an denen jeweils zwei Frauen gemeinsam teilgenommen haben. In diesem Fall entsteht nur ein einfaches Netzwerk, das durch eine relationale Variable

$$R : \Omega \times \Omega \longrightarrow \{0, 1, 2, \ldots\}$$

repräsentiert werden kann, wobei nun $R(\omega, \omega')$ die Anzahl der Ereignisse angibt, an denen ω und ω' gemeinsam teilgenommen haben. Als

[14] In dieser ambivalenten Weise wird auch in der Literatur manchmal von „sozialen Beziehungen" gesprochen, z.B. von C. Prendergast und J. D. Knottnerus (1994: 9): „A social relationship is an opportunity for social interaction, a history of shared experience, and a means of need-satisfaction."

[15] Die Daten selbst stammen aus einer früheren Untersuchung von A. Davis, B. Gardner und M. Gardner aus dem Jahr 1941.

Tabelle 2.3-1 Daten über die Teilnahme von 18 Frauen an 14 sozialen Ereignissen. Quelle: Homans (1951: 83).

		1	2	3	4	5	6	7	8	9	10	11	12	13	14
ω_1	Evelyn	x	x	x	x	x	x		x	x					
ω_2	Laura	x	x	x		x	x	x	x						
ω_3	Theresa		x	x	x	x	x	x	x	x					
ω_4	Brenda	x		x	x	x	x	x	x						
ω_5	Charlotte			x	x	x		x							
ω_6	Frances			x		x	x		x						
ω_7	Eleanor					x	x	x	x						
ω_8	Pearl						x		x	x					
ω_9	Ruth					x		x	x	x					
ω_{10}	Verne							x	x	x			x		
ω_{11}	Myra								x	x	x		x		
ω_{12}	Katherine								x	x	x		x	x	x
ω_{13}	Sylvia							x	x	x	x		x	x	x
ω_{14}	Nora						x	x		x	x	x	x	x	x
ω_{15}	Helen							x	x		x	x	x		
ω_{16}	Dorothy								x	x					
ω_{17}	Olivia									x		x			
ω_{18}	Flora									x		x			

Adjazenzmatrix erhält man dann $\mathbf{A} = \Sigma_t \mathbf{A}_t$.[16] Tabelle 2.3-2 zeigt diese Adjazenzmatrix und außerdem in der Hauptdiagonalen für jede Frau die Anzahl der Ereignisse, an denen sie teilgenommen hat.

3. Unterschiedliche Ansätze zur Definition von Gruppen. Homans hat die in Tabelle 2.3-1 angegebenen Daten verwendet, um Überlegungen zur Definition sozialer Gruppen zu illustrieren (man vgl. Homans 1951: 81ff.). Zum Verständnis ist zunächst zu beachten, dass man dieses Definitionsproblem unterschiedlich konzipieren kann:

a) Man kann sich auf (in den meisten Fällen institutionalisierte) Kriterien für die Zugehörigkeit zu einer sozialen Gruppe beziehen. Natürlich muss man sich dafür an den Auffassungen einiger oder aller der jeweils beteiligten Personen orientieren.[17]

b) Andererseits kann man versuchen, soziale Gruppen gewissermaßen aus einer Beobachterperspektive mithilfe von äußerlich feststellbaren Daten über die Interaktion von Personen zu konstruieren.

Homans verfolgt den unter (b) genannten Ansatz; daran schließen sich

[16] Erfasst man die in Tabelle 2.3-1 angegebenen Daten durch eine (18, 14)-Inzidenzmatrix $\mathbf{B}$ (wobei $b_{ij} = 1$ ist, wenn ω_i am j-ten Ereignis teilgenommen hat, und andernfalls $b_{ij} = 0$ ist), kann man die Adjazenzmatrix auch direkt berechnen: $\mathbf{A} = \mathbf{BB}'$.

[17] Kriterien für die Mitgliedschaft in einer Gruppe können von ganz unterschiedlicher Art sein und müssen nicht auf „wechselseitige positive Gefühle“ der Gruppenmitglieder Bezug nehmen, wie dies von einigen Autoren vorgeschlagen wurde; man vgl. z.B. P. V. Marsden und E. O. Laumann (1984: 58) oder L. C. Freeman (1992: 152).

Tabelle 2.3-2 Aus den Daten in Tabelle 2.3-1 berechnete Adjazenzmatrix (Anzahlen gemeinsamer Teilnahme an den sozialen Ereignissen).

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
8	6	7	6	3	4	3	3	3	2	2	2	2	2	1	2	1	1
6	7	6	6	3	4	4	2	3	2	1	1	2	2	2	1	0	0
7	6	8	6	4	4	4	3	4	3	2	2	3	3	2	2	1	1
6	6	6	7	4	4	4	2	3	2	1	1	2	2	2	1	0	0
3	3	4	4	4	2	2	0	2	1	0	0	1	1	1	0	0	0
4	4	4	4	2	4	3	2	2	1	1	1	1	1	1	1	0	0
3	4	4	4	2	3	4	2	3	2	1	1	2	2	2	1	0	0
3	2	3	2	0	2	2	3	2	2	2	2	2	2	1	2	1	1
3	3	4	3	2	2	3	2	4	3	2	2	3	2	2	2	1	1
2	2	3	2	1	1	2	2	3	4	3	3	4	3	3	2	1	1
2	1	2	1	0	1	1	2	2	3	4	4	4	3	3	2	1	1
2	1	2	1	0	1	1	2	2	3	4	6	6	5	3	2	1	1
2	2	3	2	1	1	2	2	3	4	4	6	7	6	4	2	1	1
2	2	3	2	1	1	2	2	2	3	3	5	6	8	4	1	2	2
1	2	2	2	1	1	2	1	2	3	3	3	4	4	5	1	1	1
2	1	2	1	0	1	1	2	2	2	2	2	2	1	1	2	1	1
1	0	1	0	0	0	0	1	1	1	1	1	1	2	1	1	2	2
1	0	1	0	0	0	0	1	1	1	1	1	1	2	1	1	2	2

seine Überlegungen zur Definition sozialer Gruppen an:

„We have been looking at the persons that participated together in social events. Our word for “participating together” is *interaction*: a group is defined by the interactions of its members. If we say that individuals A, B, C, D, E ... form a group, this will mean that at least the following circumstances hold. Within a given period of time, A interacts more often with B, C, D, E ... than he does with M, N, L, O, P ... whom we choose to consider outsiders or members of other groups. B also interacts more often with A, C, D, E ... than he does with outsiders, and so on for the other members of the group. It is possible just by counting interactions to map out a group quantitatively distinct from others.“ (Homans 1951: 84)

Die beiden Ansätze lassen sich mit unterschiedlichen Erkenntnisinteressen verbinden, so dass es keinen Widerspruch gibt. Im ersten Fall geht es um Gruppen, die durch ihre Wahrnehmung und normative Verankerung auch als Bedingungen für das Verhalten ihrer Mitglieder verstanden werden können. Im zweiten Fall geht es um eine empirische Beobachtung von Interaktionen, deren potentielle Relevanz als Bedingungen weiterer Interaktionen zunächst gar keine Rolle spielt.

Der zweite Ansatz führt offenbar zu der Frage, wie ausgehend von einem durch Interaktionen definierten Netzwerk Gruppen konstruiert werden können. In der Literatur, die sich mit Methoden zur Darstellung und Analyse von Netzwerken beschäftigt, sind dafür zahlreiche Verfahren vorgeschlagen worden.[18] Einige dieser Verfahren wurden auch für die von

[18] Eine umfassende Übersicht findet man bei Wasserman und Faust (1994, Kap. 7 u. 8).

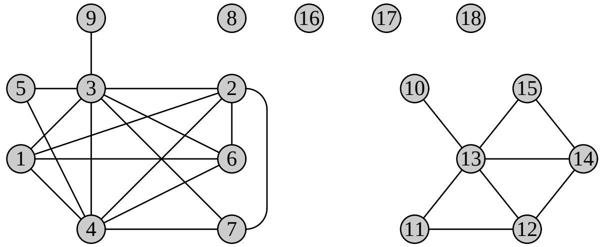
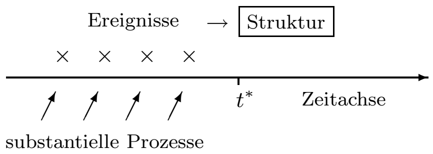


Abb. 2.3-1 Darstellung des Netzwerks mit den Daten aus Tabelle 2.3-2, wobei nur Beziehungen mit einem Wert ≥ 4 eingezeichnet sind.

Homans publizierten Daten verwendet. Zum Beispiel hat L. C. Freeman (1992) unter Bezugnahme auf M. Granovetter vorgeschlagen, zwischen „starken“ und „schwachen“ Beziehungen zu unterscheiden und Gruppen nur durch „starke“ Beziehungen zu bilden. Für die Daten in Tabelle 2.3-2 lautet sein Vorschlag, eine „starke“ Beziehung zwischen zwei Frauen dadurch zu definieren, dass sie sich mindestens viermal getroffen haben. Wie Abbildung 2.3-1 zeigt, gelangt man dann zu zwei Komponenten und vier isolierten Knoten. Natürlich entsprechen diesen Komponenten nicht unbedingt „soziale Gruppen“, die in der Wahrnehmung der beteiligten Personen handlungsrelevant sein könnten.

4. Können Strukturen als Bedingungen interpretiert werden? Das zuletzt angeführte Beispiel ist durchaus typisch dafür, wie man mithilfe formaler Methoden zu Charakterisierungen der Struktur eines Netzwerks gelangen kann. Bleiben wir bei diesem Beispiel, kann man auch feststellen, dass das, was als Struktur des Netzwerks beschrieben wird, das Ergebnis eines Prozesses ist und somit nicht als eine seiner Bedingungen verstanden werden kann. Das folgende Bild verdeutlicht diese Überlegung:



Die Struktur beschreibt einen Aspekt der Teilnahme der Frauen an den Ereignissen; sie kann also erst hinterher, etwa beginnend in einer Zeitstelle t^*, als ein Sachverhalt betrachtet werden. Und somit kann dieser Sachverhalt nicht als eine Bedingung der substantiellen Prozesse, die die Ereignisse hervorbringen, verstanden werden. Diese Überlegung bleibt offenbar auch bei einer dynamischen Betrachtung richtig, bei der man sich auf eine durch die Ereignisse erzeugte zeitliche Folge von Netzwerken bezieht (entsprechend

der Variante (a) in § 3).

Die Frage, ob und ggf. in welcher Weise Aspekte der Struktur eines Netzwerks, wie z.B. die Existenz mehrerer Komponenten, als Bedingungen irgendwelcher Prozesse (insbesondere für das Verhalten beteiligter Akteure) verstanden werden können, wird tatsächlich von den formalen Methoden zur Gruppenbildung gar nicht berührt; denn dafür müsste man von einer modalen Betrachtung der durch das Netzwerk erfassten Beziehungen ausgehen oder sich mit dem Strukturbegriff auf die substantiellen Prozesse beziehen, die die Ereignisse hervorbringen, durch die das Netzwerk definiert wird.[19]

5. Knotenzentrierte Netzwerke. Zur Beantwortung der Frage, welche Bedeutung die Struktur eines personellen Netzwerks für das Verhalten der beteiligten Akteure haben kann, ist noch eine weitere Überlegung wichtig: dass es dabei für jeden Akteur zunächst nur darauf ankommt, wie er selbst in das gesamte Netzwerk eingebunden sind. Wie diese Überlegung formal präzisiert werden kann, hängt auch von der Art der durch das Netzwerk thematisierten Beziehungen ab. Wenn deren modale Interpretation voraussetzt, dass die jeweils beteiligten Personen sich kennen und direkt kommunizieren können, erscheint es plausibel, dass für jede Person nur ihre *lokale* Einbettung in das Gesamtnetzwerk relevant ist. Um dies formal zu erfassen, dienen *knotenzentrierte Netzwerke*, die bei personellen Netzwerken auch als *ego-zentrierte Netzwerke* bezeichnet werden.

Um den Begriff, der offenbar nicht nur für personelle Netzwerke verwendet werden kann, allgemein zu definieren, beziehen wir uns auf ein Netzwerk $(\Omega, \mathcal{K})$. Für jeden Knoten $\omega \in \Omega$ kann dann ein *knotenzentriertes Netzwerk* $(\Omega_\omega, \mathcal{K}_\omega)$ definiert werden, wobei gilt: Ω_ω besteht aus ω und allen denjenigen Elementen von Ω, die mit ω durch eine Kante in $\mathcal{K}$ verbunden sind; und $\mathcal{K}_\omega$ besteht aus allen Kanten aus $\mathcal{K}$, durch die Elemente von Ω_ω verbunden werden. Bezieht man sich auf das Netzwerk in Abbildung 2.3-1, ist z.B. das knotenzentrierte Netzwerk für den Knoten Nr. 13 mit der gesamten Komponente, der dieser Knoten angehört, identisch; dagegen umfasst das knotenzentrierte Netzwerk für die Nr. 15 zusätzlich nur die Knoten 13 und 14 sowie die drei Kanten, die diese Knoten verbinden.

[19] Um diese hier zunächst nur angedeutete Möglichkeit ernsthaft zu verfolgen, müssen Modelle für substantielle Prozesse konstruiert werden.

Kapitel 3

Prozesse und Ablaufschemas

In diesem Kapitel beginnen wir mit einer Diskussion unterschiedlicher Möglichkeiten zur Konzeptualisierung von Prozessen. Im ersten Abschnitt werden einige Prozessbegriffe unterschieden und wird überlegt, wie man von historischen Prozessen und Ablaufschemas (bzw. wiederholbaren Prozessen) sprechen kann. In einem zweiten Abschnitt werden schematische Prozesskonzeptionen besprochen: Zeitreihen, insbesondere schematische Lebensverläufe, sowie statistische Prozesse, die aus zeitlichen Folgen statistischer Variablen bestehen.

3.1 Historische Prozesse und Ablaufschemas

1. Einige Varianten des Prozessbegriffs. In der Brockhaus-Enzyklopädie (Studienausgabe 2001, Bd. 17: 566) findet man als Erläuterung des Prozessbegriffs: „Verlauf, Ablauf, Hergang, Entwicklung". Der Begriff kann somit vollständig allgemein verwendet werden und setzt weder bestimmte sachliche Bezüge voraus (*was* sich verändert), noch impliziert er spezifische Annahmen über das als Prozess bezeichnete Geschehen (*wie* es sich verändert).[1] Infolgedessen ist es auch kaum möglich, eine erschöpfende

[1]Dementsprechend heißt es in einer begriffsgeschichtlichen Studie von Kurt Röttgers (1983: 93) über den heutigen Sprachgebrauch: „Alles, was sich irgendwie verändert, ist ein Prozeß oder befindet sich in einem Prozeß, so daß Ausdrücke wie 'Der Prozeß der Veränderung von X' semantisch eine bloße Verdoppelung des Ausdrucks, pragmatisch das manchmal erwünschte bloße Hinauszögern der Artikulation eines Gedankens darstellt."

Gliederung unterschiedlicher Prozesskonzeptionen vorzunehmen. Wir unterscheiden deshalb zunächst nur einige Varianten.

- Das Reden von Prozessen kann sich auf die Entwicklung von Zuständen identifizierbarer Objekte beziehen; zum Beispiel: die Entwicklung der Körpertemperatur bei einem bestimmten Menschen während eines bestimmten Zeitraums oder die Entwicklung der Niederschlagsmenge in einem bestimmten räumlichen Gebiet.
- Man kann sich auf die Entwicklung von Zuständen bei einer Mehrzahl von Objekten beziehen. Verwendet man dafür statistische Begriffsbildungen, gelangt man zu statistischen Prozessen (i.e.S.), die aus zeitlichen Folgen statistischer Variablen bestehen.
- Zu andersartigen Vorstellungen über Prozesse gelangt man, wenn man von Ereignissen ausgeht. In einer allgemeinen Formulierung erscheinen dann Prozesse als zeitlich geordnete Folgen von Ereignissen. Damit ein solcher Prozessbegriff sinnvoll verwendbar wird, muss allerdings spezifiziert werden, wie von Ereignissen gesprochen werden soll.
- Zu einer spezielleren Variante der zuletzt genannten Konzeption gelangt man, wenn man an Handlungen denkt. Prozesse erscheinen dann als Handlungszusammenhänge, die aus einer zeitlichen Abfolge einer Mehrzahl von Tätigkeiten eines oder mehrerer Akteure bestehen. Wir nennen sie *Handlungsprozesse*.

2. Prozesse werden konstruiert. Beim Reden von Prozessen muss angegeben werden, welche Arten von Veränderungen betrachtet werden sollen (*was* sich verändert); außerdem muss überlegt werden, *wie* es sich verändert bzw. verändern kann. Die zweite Frage wird uns erst später beschäftigen. Hier soll zunächst darauf aufmerksam gemacht werden, dass eine Beantwortung der ersten Frage stets eine weitgehende Selektion von Aspekten eines realen oder vorstellbaren Geschehens erfordert. Man betrachtet zum Beispiel die Entwicklung der Körpertemperatur eines Patienten und abstrahiert zugleich von beliebig vielen anderen Aspekten, die ebenfalls betrachtet werden könnten; oder man bezieht sich auf Handlungsprozesse, ohne die materiellen Kontexte, in denen sich die Handlungen abspielen, explizit in der Prozesskonzeption zu berücksichtigen.

Prozessdefinitionen beruhen also stets auf spezifischen Abstraktionen, einer Selektion bestimmter Aspekte, über deren zeitliche Entwicklung man nachdenken möchte; und alle weiteren Überlegungen beziehen sich dann ausschließlich auf den zuvor definierten Prozess, d.h. auf die jeweils ausgewählten Aspekte. So wird auch die Verwendung von Adjektiven zur Charakterisierung von Prozessen verständlich, wenn etwa von physikalischen, chemischen oder demographischen Prozessen gesprochen wird. Solche Redeweisen zeigen, dass man sich mit dem Prozessbegriff nicht unmittelbar auf Vorkommnisse und Abläufe in der menschlichen Erfahrungswelt bezieht, sondern auf Modelle, die zur Reflexion jeweils spezifischer Aspekte solcher Vorkommnisse und Abläufe und ihrer möglichen (zukünftigen) Entwicklung konstruiert werden.[2]

Diese Aussage gilt insbesondere für die an Lebensverläufen orientierte Sozialforschung. Bereits Konstruktionen individueller Biographien sind unvermeidlich sehr selektiv; und das gilt erst recht, wenn versucht wird, mithilfe statistischer Daten zu vergleichenden Aussagen über Lebensverläufe der Mitglieder einer Gesellschaft zu gelangen. Dann wird auch deutlich, dass man sich auf bestimmte Aspekte beschränken muss, aus denen Gesichtspunkte für einen Vergleich gewonnen werden können.

3. Historische Prozesse. Von *historischen Prozessen* soll in diesem Text gesprochen werden, wenn sich die Prozesskonstruktion auf einen empirisch identifizierbaren Ablauf in der menschlichen Erfahrungswelt bezieht.

– Das Reden von historischen Prozessen setzt also eine Bezugnahme auf eine menschliche Praxis voraus, durch die bzw. von der aus Prozesse identifiziert werden können.[3]

– Dem entspricht, dass zur Konzeption historischer Prozesse ein anthropozentrisches, der menschlichen Praxis gemäßes Zeitverständnis vorausgesetzt wird. Nicht nur wird vorausgesetzt, dass es zwischen Ereignissen zeitliche Beziehungen gibt, sondern außerdem eine fundamentale Unterscheidung zwischen zeitlichen Modalitäten: zwischen einer Vergangenheit, die bisher realisierte und insoweit nicht mehr veränderbare Sachverhalte umfasst, einer offenen Zukunft, die aus bisher nicht realisierten Möglichkeiten besteht, und schließlich einer flüchtigen Gegenwart, in der jeweils bestimmte Möglichkeiten realisiert und dadurch der Vergangenheit hinzugefügt werden. Offenbar verdankt sich dieses Zeitverständnis den Erfahrungen menschlicher Praxis.[4]

– Auch für historische Prozesse gilt, dass sie konstruiert werden. Es gibt indessen keine bestimmten Anforderungen an die begrifflichen Hilfsmittel. Man kann historische Prozesse als Handlungszusammenhänge oder allgemeiner als zeitliche Folgen von Ereignissen konzipieren, man kann aber auch statistische Begriffsbildungen verwenden, um historische Prozesse (z.B. die Bevölkerungsentwicklung in Deutschland während eines bestimmten Zeitraums) darzustellen. Unser Begriff historischer Prozesse soll also insbesondere keine Festlegung auf die Idee einer „Ereignisgeschichte“ beinhalten.[5]

– Unser Begriff historischer Prozesse soll auch keine Annahmen über die zeitliche Dauer voraussetzen. Christian Meier (1978: 56) hat wohl Recht: „Unter historischen Prozessen versteht man in der Regel längere, nämlich Jahrzehnte oder gar Jahrhunderte übergreifende Abläufe.“ Dem entsprechend wird bei zeitlich kürzeren Geschehnissen oft von „Ereignissen“ gesprochen. Diese Unterscheidung soll hier jedoch ausdrücklich nicht gemeint sein. Gleichwohl wird zwischen Ereignissen und Prozessen unterschieden. Wenn von Ereignissen gesprochen wird, sind zeitlich datierbare Vorkommnisse in unserer Erfahrungswelt gemeint (z.B. ein bestimmter Verkehrsunfall[6]), der Prozessbegriff bezieht sich dagegen auf theoretische Konstruktionen (beliebiger zeitlicher Dauer). Somit ist es auch möglich, ein Ereignis als einen Prozess zu betrachten, z.B. sich eine explizite Vorstellung vom Ablauf eines Verkehrsunfalls zu machen.

– Insofern die empirische Identifizierbarkeit historischer Prozesse gefordert wird, sind sie stets räumlich und zeitlich beschränkt; sie haben einen Anfang und ein (ggf. vorläufiges) Ende und gehören somit zur Vergangenheit (der Praxis, durch die sie identifiziert werden). Zur Verdeutlichung kann man an einen Spaziergang, den Entstehungsprozess

[2] Hierzu passt folgende Bemerkung von R. M. Lepsius (1976: 121): „Die Vorstellung, daß den systematischen Einzelwissenschaften jeweils abgrenzbare Teilbereiche der Erfahrung als Gegenstände ihrer Arbeit zugewiesen werden könnten, ist irrig, und insofern auch die Vorstellung, der Geschichtswissenschaft würde durch die Ausdifferenzierung der Sozialwissenschaften der Objektbereich verkleinert. Es gibt keine Erfahrungsbestände, die als solche soziologisch oder historisch sind. Erst die Umformulierung der Erfahrungsobjekte in Erkenntnisobjekte durch die Anwendung bestimmter Fragestellungen, kategorialer Bezugssysteme und Lösungswege formuliert konventionalisierte »Zuständigkeiten« von Wissenschaften.“

[3] Dem entspricht eine von A. C. Danto (1965/1980: 49) vorgeschlagene „minimale Charakterisierung der historiographischen Tätigkeit“: „daß das Unterfangen, dem Historiker sich letztendlich widmen, der Versuch ist, wahre Feststellungen über Ereignisse aus ihrer *eigenen* Vergangenheit zu treffen oder wahre Beschreibungen davon zu geben.“

[4] Ausführliche Überlegungen zu diesem anthropozentrischen Zeitverständnis findet man bei M. Oakeshott (1983: 7ff.). Es unterscheidet sich von Zeitvorstellungen, wie sie oftmals für physikalische Modelle angenommen werden und im Kontext der „New Theory of Time“ (Oaklander und Smith, 1994) diskutiert werden.

[5] Bei Historikern und Theoretikern der Geschichtswissenschaft findet man oftmals die Vorstellung, dass „Geschichte“ aus einer Folge von Ereignissen, insbesondere aus menschlichen Handlungen besteht. So spricht z.B. Gordon Leff (1969: 4) von „history“ „as the totality of human actions and endeavour“; und bei Christian Meier (1978: 11) heißt es: „Mit der Prozeß-Kategorie werden innerhalb der sozialen Welt bestimmte Handlungszusammenhänge wahrgenommen.“ Natürlich ist es zulässig, historische Prozesse so zu definieren; aber bereits innerhalb der Geschichtswissenschaft gibt es noch andere Prozesskonzeptionen, und die empirische Sozialforschung beschäftigt sich sogar überwiegend mit Prozessen, die weder unmittelbar als Handlungszusammenhänge noch ohne weiteres als „Folgen“ menschlicher Tätigkeiten konzipiert werden können.

[6] Dem entspricht folgende Bemerkung von H.-R. Jauss (1973: 554): „Ereignis ist eine objektive, für das historische Geschehen selbst konstitutive Kategorie. Das Ereignis liegt dem Zugriff des Historikers immer schon voraus; es ist nicht [im Unterschied zu unserem Prozessbegriff, G.R.] ein subjektives Schema narrativer Aneignung, sondern dessen äußere Bedingung.“ Das angeführte Beispiel soll darauf hinweisen, dass es sich bei Ereignissen auch um durchaus triviale Vorkommnisse handeln kann; dies unterscheidet den allgemeinen von einem emphatischen Ereignisbegriff, den Christian Meier (a.a.O., S. 47) so erläutert: „Als Ereignis bezeichnen wir im Alltag primär ein besonderes, aus dem Üblichen herausragendes Geschehen. Die Historie gebraucht das Wort im gleichen Sinne und meint damit zumeist die bemerkenswerten, »denk- (und überlieferungs-) würdigen« Handlungen und Handlungszusammenhänge sowie anderswie bewirkten Einschnitte des politischen und militärischen Bühnengeschehens.“

eines Gebäudes oder an die Bevölkerungsentwicklung in einer bestimmten Region während eines bestimmten Zeitraums denken. Offenbar kann man beliebig viele historische Prozesse dieser Art konzipieren. Oftmals lassen sich auch Beziehungen zwischen mehreren Prozessen herstellen. Es ist jedoch fragwürdig, ob man sinnvoll von einer Gesamtheit aller historischen Prozesse, die in bestimmter Weise miteinander verbunden sind, sprechen kann.

4. Ablaufschemas. Insofern bei der Konzeption historischer Prozesse ein anthropozentrisches Zeitverständnis vorausgesetzt wird, sind sie „einmalig". Unabhängig von dieser Feststellung kann man jedoch von *wiederholbaren Prozessen* sprechen, d.h. von Prozessen, die in ähnlicher Form mehrmals ablaufen können. Wiederholbarkeit in diesem Sinn setzt nur voraus, dass mehrere Prozesse unter bestimmten Aspekten als vergleichbar betrachtet werden können.

Bei wiederholbaren Prozessen ist offenbar eine begriffliche Unterscheidung erforderlich: Einerseits kann man sich auf die Form des Prozessablaufs beziehen, wir nennen dies ein *Ablaufschema* (eines wiederholbaren Prozesses); andererseits kann man sich auf jeweils individuelle Prozessabläufe beziehen, die als Realisationen des Ablaufschemas betrachtet werden können. Sofern diese Realisationen nicht nur fiktiv vorgestellt werden, sondern tatsächlich stattfinden, handelt es sich um historische Prozesse. Somit gibt es auch keinen begrifflichen Gegensatz zwischen historischen und wiederholbaren Prozessen. Als Gegensatz zu historischen Prozessen kann man an fiktive Prozesse denken, die man sich nur vorstellt. Wenn man einen historischen (oder fiktiven) Prozess wiederholbar nennt, ist dagegen gemeint, dass man ihn als Realisation eines Ablaufschemas betrachten möchte und dass es noch andere Realisationen dieses Ablaufschemas gibt oder geben kann.[7] Einige Beispiele können das verdeutlichen.

– Eine wichtige Klasse von Beispielen liefern Computerprogramme (oder abstrakter: Algorithmen). Bei einem Computerprogramm muss man offenbar unterscheiden zwischen einerseits dem Programm, das in diesem Fall das Ablaufschema bildet und die möglichen Programmabläufe festlegt, und andererseits den Prozessrealisationen, also den Programmabläufen, die stattfinden, wenn das Programm gestartet wird. Jeder Programmablauf ist seinerseits ein historischer Prozess, der in einem bestimmten räumlichen und zeitlichen Kontext stattfindet.[8]
– Weitere leicht durchschaubare Beispiele liefern Gesellschaftsspiele wie z.B. Schach oder Skat. Einerseits gibt es Spielregeln, durch die festgelegt wird, wie Spiele ablaufen können; andererseits gibt es jeweils bestimmte Spielabläufe, die man sich vorstellen oder als historische Prozesse realisieren kann.

[7] *Wiederholbar zu sein* ist also keine Eigenschaft, die einem Prozess „an und für sich" zukommt, sie resultiert vielmehr aus einer jeweils bestimmten Betrachtungsweise eines historischen oder fiktiven Prozesses.

[8] Man vgl. dazu auch die Überlegungen von B. C. Smith (1996: 32ff.).

– Weiterhin kann man an viele andere Handlungsprozesse denken, die mehr oder weniger detailliert durch Regeln bestimmt werden und für die es insofern ein Ablaufschema gibt, zum Beispiel die Zubereitung von Speisen (nach einem Rezept) oder die Durchführung einer medizinischen Diagnose (nach den dafür gültigen ärztlichen Regeln).
– Schließlich kann man auch menschliche Lebensverläufe als wiederholbare Prozesse auffassen. Zwar kann niemand das eigene Leben wiederholen, aber die Idee der Wiederholbarkeit bezieht sich nur darauf, dass man die Lebensverläufe mehrerer Menschen (aller Mitglieder einer Gesellschaft) vergleichend betrachten und zu diesem Zweck Ablaufschemas konstruieren kann. Eine Möglichkeit liefern sogenannte Biographieschemas, die im nächsten Abschnitt besprochen werden.

Die Beispiele zeigen, dass wiederholbare Prozesse von ganz unterschiedlicher Art sein können. Während Computerprogramme Beispiele für mechanische Prozesse sind, die durch Akteure nur initialisiert werden,[9] handelt es sich bei Gesellschaftsspielen um Handlungsprozesse, bei denen auch der Ablauf durch Akteure beeinflusst wird. Schließlich können Lebensverläufe weder als rein mechanische Prozesse noch ausschließlich als Handlungsprozesse begriffen werden.

Weiterhin zeigen die Beispiele, dass es bei Ablaufschemas unterschiedliche *Spielräume für Prozessrealisationen* geben kann. Einen Extremfall bilden Algorithmen, die einen Prozess eindeutig festlegen. Die meisten Ablaufschemas lassen dagegen in mehr oder weniger weiten Grenzen unterschiedliche Prozessabläufe zu.[10]

5. Durch Regeln bestimmte Prozesse. In vielen Fällen können Ablaufschemas für wiederholbare Prozesse durch Regeln beschrieben werden. Man kann dann davon sprechen, dass die Prozessrealisationen „durch Regeln bestimmt" werden. Allerdings muss darauf geachtet werden, was mit dieser Formulierung ausgesagt werden kann.

[9] Wir sprechen in diesem Text von *mechanischen* Prozessen, wenn Akteure gar nicht beteiligt sind oder nur als Auslöser des Prozesses, also ohne Einfluss darauf zu nehmen, wie der Prozess abläuft. (Die obige Aussage bezieht sich somit nicht auf sogenannte interaktive Programme, bei denen ein Akteur in den Programmablauf eingreifen kann.) In dieser Bedeutung bezieht sich das Adjektiv 'mechanisch' also nicht auf die Mechanik als Teilgebiet der Physik, sondern dient der Charakterisierung einer bestimmten Art von Prozessen.

[10] Der Begriff eines Ablaufschemas soll also nicht beinhalten, dass nur ein in irgendeinem Sinn „kleiner" Spielraum für dem Schema entsprechende Prozesse besteht. Deshalb kann auch nicht bereits durch den Verweis auf ein Ablaufschema von „typischen" Prozessabläufen gesprochen werden; obwohl – wie etwa von K.-G. Faber (1971: 95) ausgeführt worden ist – der Typenbegriff sich ebenfalls bloß in abstrakter Weise auf reales bzw. realisierbares Geschehen bezieht.

Zunächst ist klar: Dass ein Prozess durch Regeln bestimmt wird, impliziert nicht, dass der Prozessablauf determiniert ist. Gilbert Ryle (1949/1982: 98ff.) hat das am Beispiel des Schachspiels verdeutlicht: obwohl durch Regeln bestimmt, ist jeder einzelne Spielverlauf nicht (jedenfalls nicht durch die Spielregeln) determiniert. Das Beispiel zeigt auch, dass Regeln, die einen Prozess bestimmen, nicht als Ursachen verstanden werden können, die gewissermaßen bewirken, dass sich der Prozessablauf an die Regeln hält (und zwar gilt diese Feststellung ganz unabhängig von der Größe des durch die Regeln gegebenen Spielraums für Prozessrealisationen): Die Regeln, die es für Schachspiele gibt, können weder bewirken noch garantieren, dass sich die Spieler an die Regeln halten.

Dies gilt aber auch für mechanische Prozesse, an deren Ablauf keine Akteure beteiligt sind. Zum Beispiel kann auch ein Computerprogramm weder bewirken noch garantieren, dass bei seiner Aktivierung ein dem festgelegten Ablaufschema entsprechender Prozess stattfindet. Wie der historische Prozess abläuft, der bei der Programmaktivierung beginnt, hängt vielmehr von zahlreichen Bedingungen ab, die durch das Computerprogramm überhaupt nicht in Betracht gezogen werden (z.B. von der Stromversorgung des Prozessors).

Allerdings ist es durchaus von Bedeutung, ob Akteure einen Prozess nur initialisieren oder ob es auch von ihrem Verhalten abhängt, wie der Prozess abläuft (wie dies inbesondere, aber nicht nur bei Handlungsprozessen der Fall ist). Denn wenn letzteres der Fall ist, kann man in zwei unterschiedlichen Bedeutungen von der Existenz von Regeln sprechen.

- Zunächst kann gemeint sein, dass sich die Regeln auf ein Modell beziehen, dass *aus einer Beobachterperspektive* konstruiert wird, um wiederholbare Prozessabläufe vorstellbar und reflektierbar zu machen. Bei mechanischen Prozessen ist offenbar nur dieses Verständnis von Regeln möglich.
- Wenn es sich jedoch um Prozesse handelt, an deren Ablauf Akteure beteiligt sind, wird es möglich und ist oftmals der Fall, dass auch diese Akteure über ein Regelwissen verfügen. Es gibt dann ein Regelwissen *aus der Akteursperspektive*, und wie ein Prozess abläuft hängt auch davon ab, wie dieses Regelwissen beschaffen ist und wie die beteiligten Akteure es in ihrem Verhalten verwenden.

Oftmals kann aus beiden Perspektiven das gleiche (oder zumindest ein teilweise gleiches) Regelwissen unterstellt werden, z.B. bei Gesellschaftsspielen, bei denen Spieler und Zuschauer die Regeln gleichermaßen kennen. Anders verhält es sich jedoch in vielen Fällen, in denen Wissenschaftler Modelle für Prozesse konstruieren, an denen Menschen beteiligt sind. Dies geschieht normalerweise aus einer Beobachterperspektive, und es kann meistens nicht angenommen werden, dass das wissenschaftlich konstruierte Regelwissen auch bei den Akteuren der Prozesse vorhanden ist (z.B. bei einem Modell zur Erklärung des Zustandekommens von Staus auf Autobahnen). Dementsprechend muss auch zwischen Verhaltensregelmäßigkeiten, die aus einer Beobachterperspektive festgestellt werden können, und Regeln, an denen sich Akteure in ihrem Verhalten orientieren, begrifflich unterschieden werden.

3.2 Zeitreihen und statistische Prozesse

Wenn man sich im Rahmen der empirischen Sozialforschung mit Prozessen beschäftigt, werden diese fast immer als Realisationen theoretisch konzipierter Ablaufschemas betrachtet. Möglichkeiten zur Konstruktion von Ablaufschemas hängen in erster Linie von der Konzeption der Prozesse ab. In diesem Abschnitt beziehen wir uns auf Zeitreihen und statistische Prozesse.

1. Zeitachsen. Offenbar benötigt man zur Konzeption von Ablaufschemas einen zeitlichen Rahmen. Meistens wird eine Zeitachse verwendet. Es gibt hauptsächlich zwei Varianten:

- Man kann sich die Zeit als eine *Folge von Zeitstellen* (z.B. Sekunden, Stunden, Tage, Monate, Jahre) vorstellen. Zur Repräsentation der Zeitstellen werden die natürlichen oder ganzen Zahlen verwendet, und man spricht von einer *diskreten Zeitachse*.
- Man kann versuchen, sich die Zeit als ein linear geordnetes Kontinuum von Zeitpunkten vorzustellen. Zur Repräsentation werden in diesem Fall die reellen Zahlen verwendet, und man spricht von einer *stetigen* oder *kontinuierlichen Zeitachse*.

Unabhängig von dieser Unterscheidung, die die begriffliche Repräsentation von Zeit(stellen) betrifft, kann man Verwendungskontexte unterscheiden. Zunächst kann man an eine *historische Zeitachse* denken, die zur Repräsentation der historischen Zeit dient, in der sich das Leben der Menschen tatsächlich abspielt. Gedankliche Bezugnahmen auf diese Zeitachse erfolgen mithilfe von Kalendern und Uhren.[11] Davon zu unterscheiden sind *Modell-Zeitachsen* (auch *Prozesszeitachsen* genannt), die zur Konstruktion von Modellen verwendet werden, die einer modalen Reflexion von Prozessabläufen dienen sollen.

Will man eine explizite Repräsentation zeitlicher Bezüge vornehmen, muss man sich für eine diskrete oder eine stetige Darstellung entscheiden. Wir werden in diesem Text in den meisten Fällen eine diskrete Zeitachse zugrunde legen und dafür die Notation

$$\mathcal{T} := \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$$

[11] Darüber, wie sich solche Orientierungsmittel historisch entwickelt und verändert haben, gibt es eine umfangreiche Literatur, man vgl. z.B. E. G. Richards (1998).

verwenden, so dass $\mathcal{T}$ formal der Menge der ganzen Zahlen entspricht und die Ordnungsrelation zwischen diesen Zahlen ($\leq$ und $\geq$) als zeitliche Relation zwischen den durch sie bezeichneten Zeitstellen verstanden werden kann. Festlegungen über die Art der Zeitstellen und Verknüpfungen mit einer historischen Zeitachse können bei Bedarf erfolgen. Natürlich benötigt man zur Repräsentation von Daten oft nur einen Teil der Zeitachse; wir verwenden dann die Notation $\mathcal{T}^*$, womit stets eine zusammenhängende und meistens (wenn nicht ausdrücklich anders angegeben) auch endliche Teilmenge von $\mathcal{T}$ gemeint sein soll.

2. Zeitreihen. Sobald man über den Begriff einer Zeitachse verfügt, kann man in sehr allgemeiner Weise von *Zeitreihen* sprechen. Zur formalen Vergegenwärtigung kann ein *Zeitreihenschema*

$$X : \mathcal{T}^* \longrightarrow \tilde{\mathcal{X}} \tag{3.1}$$

verwendet werden, also eine Funktion, durch die jeder Zeitstelle t einer Zeitachse $\mathcal{T}^*$ ein Wert $X(t)$ in einem Wertebereich $\tilde{\mathcal{X}}$ zugeordnet wird. Je nachdem, ob es sich bei $\mathcal{T}^*$ um eine kontinuierliche oder diskrete Zeitachse handelt, kann man kontinuierliche (stetige) und diskrete Zeitreihen unterscheiden. Hinsichtlich des Wertebereichs $\tilde{\mathcal{X}}$ kann man außerdem folgende Unterscheidungen treffen:

- *Einfache Zeitreihen.* In diesem Fall wird jeder Zeitstelle t ein einfacher Wert $X(t)$ zugeordnet, so dass $\tilde{\mathcal{X}}$ numerisch durch reelle Zahlen repräsentiert werden kann; zum Beispiel: die Entwicklung der Körpertemperatur eines Patienten oder der Bevölkerungszahl eines Landes während eines gewissen Zeitraums.
- *Vektorielle Zeitreihen.* In diesem Fall wird jeder Zeitstelle t ein Vektor $X(t) = (X_1(t), \ldots, X_m(t))$ zugeordnet. Als Beispiel kann man daran denken, dass bei neugeborenen Kindern für einen gewissen Zeitraum sowohl die Körpergröße als auch das Körpergewicht erfasst wird.
- *Funktionale Zeitreihen.* In diesem Fall wird jeder Zeitstelle t eine Funktion zugeordnet; $X(t)$ ist dann keine Zahl, sondern eine Funktion (und wir verwenden dann meistens die Schreibweise X_t anstelle von $X(t)$). Bei diesen Funktionen kann es sich insbesondere um statistische Variablen handeln, so dass statistische Prozesse entstehen (das wird weiter unten genauer besprochen).

Bei einfachen und vektoriellen Zeitreihen kann der Wertebereich $\tilde{\mathcal{X}}$ dem Merkmalsraum einer ein- bzw. mehrdimensionalen statistischen Variablen entsprechen (man vgl. die Ausführungen in Abschnitt 1.1). Auch dann ist jedoch die Analogie zwischen dem Zeitreihenschema (3.1) und dem für statistische Variablen verwendeten Schema $X : \Omega \longrightarrow \tilde{\mathcal{X}}$ rein formal, denn eine Zeitachse $\mathcal{T}^*$ ist keine Objektmenge. Tatsächlich wird in der allgemeinen Definition von Zeitreihen überhaupt kein bestimmter Objektbezug

vorgenommen; insofern handelt es sich um eine rein formale Definition, deren inhaltliche Bedeutung sich nur aus einem Anwendungskontext ergeben kann.

3. Schematische Lebensverläufe. Die Methodenliteratur befasst sich überwiegend mit Zeitreihen, deren Wertebereiche quantitativ, meistens auch metrisch sind. In der empirischen Sozialforschung sind auch qualitative Zeitreihen wichtig, da sie sich zur Konzeptualisierung schematischer Lebensverläufe eignen. Dabei kann man grundsätzlich an beliebige Objekte, insbesondere auch an Menschen denken. Die allgemeine Vorstellung besteht darin, dass sich ein Objekt im Zeitablauf in unterschiedlichen Zuständen befinden kann. Um dies zu erfassen, eignet sich das allgemeine Zeitreihenschema $Y : \mathcal{T}^* \longrightarrow \tilde{\mathcal{Y}}$, wobei jetzt $\tilde{\mathcal{Y}}$ ein *Zustandsraum* ist, dessen Elemente Zustände sind, in denen sich Objekte einer bestimmten Art während der durch $\mathcal{T}^*$ gegebenen Zeitstellen befinden können. Eine Zeitreihe $\{Y(t) \,|\, t \in \mathcal{T}^*\}$ erfasst dann aspekthaft einen Ausschnitt eines Lebensverlaufs. Die Formulierung ist allgemein genug, um sich sowohl auf zeitliche Ausschnitte eines Lebensverlaufs als auch auf die gesamte Lebensdauer zu beziehen.

Wiederum muss unterschieden werden zwischen spezifischen Lebensverläufen (Zeitreihen), die für ein jeweils spezifisches Objekt erfasst werden, und dem zugrundeliegenden Zeitreihenschema, das ein Ablaufschema für *mögliche* Lebensverläufe liefert. In diesem Zusammenhang wird das Ablaufschema auch als ein *Biographieschema* bezeichnet (Rohwer und Pötter 2001: 186), dessen Definition zweierlei erfordert:

- Zunächst muss ein Zustandsraum $\tilde{\mathcal{Y}}$ festgelegt werden, so dass dem Objekt in jeder Zeitstelle $t \in \mathcal{T}^*$ genau ein Zustand zukommt (natürlich kann auch ein mehrdimensionaler Zustandsraum verwendet werden, der eine simultane Bezugnahme auf mehrere Zustände erlaubt);
- außerdem muss für jeden Zustand festgelegt werden, in welche Folgezustände ein Wechsel stattfinden kann.

Um die Begriffsbildung zu illustrieren, zeigt Abbildung 3.2-2 ein Biographieschema für die Bildung und Auflösung von Lebensgemeinschaften. Es gibt folgende Zustände: (A) Anfangszustand, in dem sich eine Person vor der ersten Lebensgemeinschaft befindet, (U) nicht-eheliche Lebensgemeinschaft, (H) eheliche Lebensgemeinschaft (verheiratet), (S) nach einer Trennung bzw. Scheidung, (E) Endzustand (Tod). Die Pfeile deuten die möglichen Zustandsveränderungen an. Zum Beispiel gibt es keinen Pfeil von A nach S, weil eine Trennung oder Scheidung das Bestehen einer Lebensgemeinschaft voraussetzt.

4. Ein spezieller Ereignisbegriff. Mit dem allgemeinen Zeitreihenschema (3.1) ist ein spezieller Ereignisbegriff verbunden, durch den Ereignisse *als Zustandswechsel* aufgefasst werden. Wenn $\mathcal{T}^*$ eine diskrete Zeitachse ist, kann man mit diesem Begriff von einem Ereignis im Übergang von der

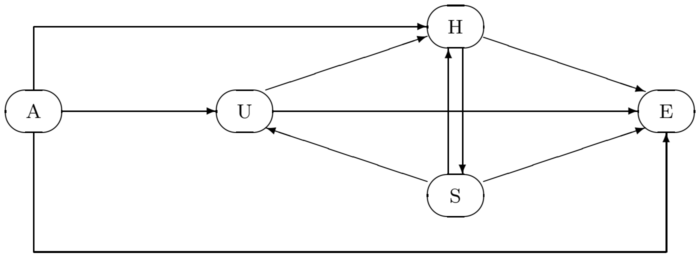


Abb. 3.2-2 Ein Biographieschema für die Bildung und Auflösung von Lebensgemeinschaften (Rohwer und Pötter 2001: 186).

Zeitstelle t zur nächsten Zeitstelle $t+1$ sprechen, wenn $X(t+1) \neq X(t)$ ist. Ein so definiertes Ereignis kann also durch die Angabe einer Zeitstelle t und der beiden Zustände $X(t)$ und $X(t+1)$ vollständig charakterisiert werden (wobei angenommen wird, dass die ontologischen Bezüge durch Erläuterungen des vorausgesetzten Zeitreihenschemas gegeben sind).

Dieser Ereignisbegriff wird insbesondere in der Literatur, die sich mit statistischen Methoden zur Modellierung qualitativer Zeitreihen beschäftigt, oft verwendet.[12] Es handelt sich jedoch um einen durchaus speziellen Ereignisbegriff.

- Tatsächlich können die meisten Ereignisse nicht hinreichend durch Zustandswechsel beteiligter Objekte beschrieben werden. Vielmehr verweist der Ereignisbegriff in seiner normalen Verwendung zunächst auf ein Geschehen, an dem oft Tätigkeiten von Akteuren beteiligt sind, die sich nicht durch einen Zustandswechsel charakterisieren lassen (man denke z.B. an einen Verkehrsunfall).
- Hier muss auch daran erinnert werden, dass das Zeitreihenschema, auf das sich der spezielle Ereignisbegriff bezieht, selbst bereits eine durchaus spezifische Konzeption von Prozessen als *zeitliche Folgen von Zuständen* impliziert. Insbesondere können Handlungsprozesse durch dieses Schema nicht erfasst werden.
- Schließlich ist bemerkenswert, dass mit dem speziellen Ereignisbegriff der direkte Realitätsbezug verschwindet, der mit dem normalen Reden von Ereignissen verbunden ist. Wie bereits erwähnt wurde (Anm. 6 auf S. 53), bezieht man sich mit dem Ereignisbegriff normalerweise auf Geschehnisse in der menschlichen Erfahrungswelt. Der spezielle Ereignisbegriff bezieht sich dagegen auf Prozesse, die als Modelle eines realen oder vorstellbaren Geschehens konzipiert werden. Dann ist z.B. mit einer Heirat nicht das reale Ereignis gemeint, das in einer bestimmten

[12] Dementsprechend wird auch von „Techniques of Event History Modeling“ (Blossfeld und Rohwer 2002) gesprochen.

Weise stattgefunden hat, sondern ein daraus ableitbarer Zustandswechsel bei einer Person, die hinterher verheiratet ist.

5. Zeitliche Folgen statistischer Variablen. Das allgemeine Zeitreihenschema (3.1) umfasst insbesondere *statistische Prozesse*, worunter in diesem Text zeitliche Folgen statistischer Variablen verstanden werden sollen. Wir unterscheiden drei Varianten, wobei der Zeitindex t jeweils Werte in einer diskreten Zeitachse $\mathcal{T}^*$ annehmen kann.

a) Wir sprechen von einem *synchron aggregierten* statistischen Prozess, wenn es für jede Zeitstelle t eine Objektmenge Ω_t gibt, so dass der Prozess aus einer Folge von Variablen besteht:

$$X_t : \Omega_t \longrightarrow \tilde{\mathcal{X}} \tag{3.2}$$

b) Wir sprechen von einem *diachron aggregierten* statistischen Prozess, wenn es nur eine Objektmenge Ω gibt, deren Elementen jedoch für alle Zeitstellen bestimmte Merkmalswerte zugeordnet werden können.[13] Der Prozess besteht somit aus einer Folge von Variablen

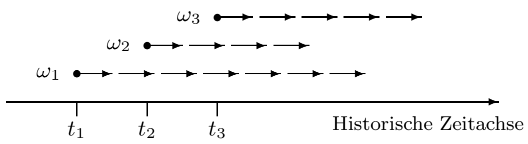
$$X_t : \Omega \longrightarrow \tilde{\mathcal{X}} \tag{3.3}$$

c) Schließlich gibt es die Möglichkeit, dass sich die Zugehörigkeit von Elementen zur Objektmenge im Zeitablauf verändern kann und Informationen darüber bei der Prozesskonstruktion berücksichtigt werden können (so dass anders als im Fall (a) Elemente während ihrer Mitgliedschaft in den Objektmengen identifizierbar bleiben). Wir sprechen dann von *transitorisch aggregierten* statistischen Prozesse. Ein wichtiges Beispiel bilden demographische Prozesse, bei denen sich eine Gesellschaft durch Geburten, Sterbefälle und Migrationen verändert

6. Längsschnittgesamtheiten und Prozesszeitachsen. Offenbar können sich bei synchron aggregierten statistischen Prozessen die Objektmengen Ω_t in jeder Zeitstelle verändern. Dagegen geht man bei diachron aggregierten statistischen Prozessen von einer *Längsschnittgesamtheit* Ω aus, die für die Prozessdefinition als unveränderlich angenommen wird. Man kann sich vorstellen, dass jedem Element von Ω zunächst ein individueller Prozess zugeordnet ist und dass der statistische Prozess aus einer Aggregation dieser individuellen Prozesse entsteht.

Natürlich verlaufen die individuellen Prozesse nicht unbedingt zeitlich parallel, und sie haben auch meistens unterschiedliche zeitliche Dauern, wie durch folgendes Bild illustriert wird:

[13] Diese Formulierung soll so verstanden werden, dass ggf. auch eine Kennzeichnung verwendet werden kann, die besagt, dass ein Objekt in einer Zeitstelle noch nicht oder nicht mehr existiert.



In diesem Bild wird angenommen, dass die individuellen Prozesse bei drei Objekten in drei verschiedenen Zeitstellen einer historischen Zeitachse beginnen. Bezieht sich die Prozessdefinition z.B. auf Ehedauern, kann man sich vorstellen, dass es sich um drei Personen mit einem unterschiedlichen Heiratsdatum handelt. Das Bild macht auch deutlich, dass es zur Bildung einer Längsschnittgesamtheit drei Möglichkeiten gibt:

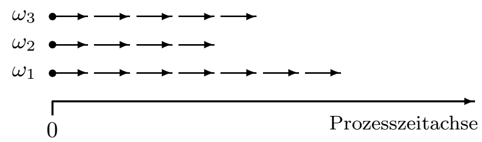
a) Man kann Objekte zusammenfassen, bei denen der Prozess in der gleichen Zeitstelle beginnt (z.B. Personen, die im gleichen Jahr geheiratet haben, wenn man als Zeitstellen Jahre verwendet);
b) man kann Objekte zusammenfassen, bei denen der Prozess in der gleichen Zeitstelle aufhört (z.B. Personen, die im gleichen Jahr geschieden worden sind); und
c) man kann Objekte zusammenfassen, bei denen der Prozess irgendwann während eines längeren Zeitraums begonnen (oder aufgehört) hat.

In der Literatur wird meistens die erste Variante empfohlen; man spricht dann von einem *Kohortenansatz*, wobei unter einer *Kohorte* eine Gesamtheit von Menschen verstanden wird, bei denen ein Prozess (einer bestimmten Art) in der gleichen historischen Zeitstelle begonnen hat. Allerdings gibt es keine vollständig scharfe Abgrenzung zum Fall (c), denn man kann auch von längeren Zeitperioden ausgehen, in denen das eine Kohorte definierende Ereignis stattgefunden hat.[14] Immerhin gibt es einen deutlichen Unterschied zum Fall (b), insbesondere wenn man von individuellen Prozessen ausgeht, die sich in ihrer zeitlichen Dauer erheblich unterscheiden können.

Ein Kohortenansatz wird in der an Lebensverläufen orientierten Sozialforschung in erster Linie als ein Hilfsmittel zur Darstellung historischer Veränderungen in der Entwicklung von Lebensverläufen verwendet. Unabhängig davon, wie eine Längsschnittgesamtheit gebildet wird, geht es jedoch stets um einen statistischen Vergleich der für die Elemente der Gesamtheit erfassten individuellen Prozesse. Dafür wird in jedem Fall eine *Prozesszeitachse* verwendet, so dass man sich vorstellen kann, dass alle individuellen Prozesse in der gleichen Prozesszeitstelle beginnen und somit während ihrer Dauer synchron ablaufen. In unserem Beispiel kann der Übergang von einer historischen zu einer Prozesszeitachse folgendermaßen

[14] Dem entspricht z.B. folgende Definition von N. D. Glenn (1977: 8): „a cohort is defined as those people within a geographically or otherwise delineated population who experienced the same significant life event within a given period of time.“

veranschaulicht werden:



Es wird also angenommen, dass bei jedem Objekt der Prozess in einer Zeitstellt $t = 0$ beginnt. Dementsprechend wird in diesem Text zur numerischen Repräsentation einer diskreten Prozesszeitachse die Notation $\mathcal{T}_0 := \{0, 1, 2, 3, \ldots\}$ verwendet.

7. Individuelle und aggregierte Prozesse. Zur Erläuterung dieser Unterscheidung beziehen wir uns auf das Schema (3.3) für diachron aggregierte statistische Prozesse.

- Einerseits kann man sich dann bei jedem Objekt $\omega \in \Omega$ auf einen diesem Objekt individuell zurechenbaren Prozess $\{X_t(\omega) \,|\, t \in \mathcal{T}^*\}$ beziehen.[15] Offenbar kann ein solcher Prozess auch als eine Zeitreihe bzw. als Realisation eines Zeitreihenschemas aufgefasst werden.
- Andererseits kann man sich auf einen *aggregierten Prozess* beziehen, der aus der statistischen Aggregation der individuellen Prozesse entsteht und durch die gemeinsame Verteilung der Variablen X_t oder daraus ableitbare Zeitreihen charakterisiert werden kann.

Daraus ergibt sich zugleich ein wichtiger Unterschied zu synchron aggregierten statistischen Prozessen, bei denen man sich nur auf eine zeitliche Folge von Verteilungen der Variablen X_t, nicht jedoch auf ihre gemeinsame Verteilung beziehen kann.[16] Ein weiterer bemerkenswerter Unterschied besteht darin, dass es in diesem Fall nicht möglich ist, sich in gedanklich bestimmter Weise auf korrespondierende individuelle Prozesse zu beziehen. Denn entweder, wie im Ansatz des Schemas (3.2), wird bereits bei der Prozessdefinition gar nicht von individuellen Prozessen ausgegangen; oder es wird zwar von individuellen Prozessen ausgegangen, dann jedoch von der diachronen Identität der diese individuellen Prozesse konstituierenden Objekte abstrahiert.

Man kann sich leicht klarmachen, dass mit dieser Abstraktion ein erheblicher Informationsverlust verbunden sein kann. Zur Illustration kann

[15] Bei dieser Schreibweise wird vorausgesetzt, dass es für alle Zeitstellen zumindest formal zurechenbare Zustände gibt. Natürlich ist es möglich, die Definition jedes individuellen Prozesses auf diejenigen Zeitstellen einzuschränken, in denen das betreffende Objekt existiert.

[16] Es sei angemerkt, dass solche zeitlichen Folgen von Verteilungen bei Rohwer und Pötter (2001: 192) missverständlich als „diachrone Zustandsverteilungen“ bezeichnet werden. Besser sollte von zeitlichen Folgen synchroner (Zustands-)Verteilungen gesprochen werden.

bereits eine Gesamtheit dienen, die nur aus zwei Personen (ω_1 und ω_2) besteht, die sich in zwei verschiedenenen Zuständen (etwa 0 = erwerbstätig und 1 = arbeitslos) befinden können. Dann können zwei Prozessvarianten etwa folgendermaßen aussehen:

t	0	1	2	3	4	5	6
ω_1	1	0	1	0	1	0	1
ω_2	0	1	0	1	0	1	0

t	0	1	2	3	4	5	6
ω_1	0	0	0	0	0	0	0
ω_2	1	1	1	1	1	1	1

Offenbar unterscheiden sich die beiden Varianten, denn im ersten Fall sind beide Personen abwechselnd erwerbstätig und arbeitslos, wohingegen im zweiten Fall eine Person nie, die andere Person immer arbeitslos ist. Dennoch würde bei einer synchronen Aggregation in beiden Fällen der gleiche statistische Prozess enstehen.

Kapitel 4

Regeln und dynamische Modelle

In diesem Kapitel beginnen wir mit einer Diskussion von Modellen, die sich auf modale Fragestellungen beziehen. Um einige Grundgedanken zu erläutern, werden dynamische Modelle besprochen, die an das in Abschnitt 3.2 definierte allgemeine Zeitreihenschema anknüpfen. Es sei aber betont, dass die in der Sozialforschung verwendeten Modelle nicht unbedingt eine explizite Bezugnahme auf eine Zeitachse erfordern; und im nächsten Kapitel werden deterministische und stochastische Varianten funktionaler Modelle diskutiert, die keinen expliziten Zeitbezug aufweisen.

Unabhängig davon, ob es einen expliziten Zeitbezug gibt, beziehen sich die in diesem Text betrachteten Modelle auf „durch Regeln bestimmte Prozesse“ (vgl. Abschnitt 3.1, § 5). Wir beginnen deshalb mit einer Diskussion unterschiedlicher Verwendungsweisen des Regelbegriffs. Im zweiten Abschnitt werden Ansätze zur Konstruktion von Modellen für Prozesse dargestellt, deren Konzeption an das in Abschnitt 3.2 definierte allgemeine Zeitreihenschema anknüpft. Im dritten Abschnitt werden stochastische Ablaufschemas und Prozesse besprochen.

4.1 Verwendungen des Regelbegriffs

1. Eingrenzung des Regelbegriffs. Die Brockhaus-Enzyklopädie (Studienausgabe 2001, Band 18: 159) gibt folgende Erläuterung: „Regel [aus mlat. regula >Ordensregel<, von lat. regula >Richtholz<, >Richtschnur<; >Regel<], 1) *allg.*: Richtlinie, Norm, Vorschrift.“ Weitere Erläuterungen betreffen spezielle Wortverwendungen, von denen hier abgesehen werden kann. Bemerkenswert ist, dass es viele weitere Worte mit einer ähnlichen Bedeutung gibt: Anweisung, Rezept, Verfahren, Methode, Maxime, Kanon, Gesetz; außerdem kann das Wort in zahlreichen Wortverbindungen verwendet werden, z.B. kann man von Spielregeln, Ordensregeln und Verkehrsregeln sprechen. Es erscheint aussichtslos, eine Wortbedeutung zu fixieren, die allen Verwendungsmöglichkeiten gerecht werden kann. In diesem Text orientiere ich mich an folgenden Überlegungen zum Wortgebrauch:

– Regeln können als *gedankliche Hilfsmittel zur Orientierung im Denken und Handeln* verstanden werden;[1] also als Hilfsmittel für menschliche Tätigkeiten, aber als *gedankliche* Hilfsmittel im Unterschied zu materiellen Hilfsmitteln wie z.B. Straßen und Computern.

– Als Hilfsmittel für Tätigkeiten sind Regeln von den Tätigkeiten, in denen Menschen Regeln verwenden – ihnen folgen, sich an ihnen orientieren, mit ihrer Hilfe Tätigkeiten strukturieren oder argumentieren oder Erwartungen bilden –, zu unterscheiden. Zum Beispiel ist ein Kochrezept von den Tätigkeiten zu unterscheiden, in denen man dem Rezept entsprechend ein Gericht zubereitet. (Dies ist einer der Gründe dafür, dass die Darstellung einer Regel nicht mit einer Beschreibung von Tätigkeiten identifiziert werden kann; gleichwohl kann es natürlich zur Erläuterung einer Regel oft sinnvoll sein, auf entsprechende Tätigkeiten hinzuweisen.)

– Regeln haben (meistens) unbestimmt viele Anwendungsfälle. Die Aufforderung, in einer bestimmten Situation etwas Bestimmtes zu tun, ist deshalb keine Regel; wohl aber könnte eine Regel in der Anweisung bestehen, dass immer dann, wenn eine Situation einer bestimmten Art vorliegt, etwas Bestimmtes getan werden soll.[2]

– Dass es eine Regel gibt, soll zunächst nur bedeuten, dass sie formuliert werden kann und dass vorstellbar ist, dass sie verwendet werden könnte. Der Begriff einer Regel impliziert somit keine Festlegungen oder Annahmen über ihre Geltung (so dass auch die Frage, was die Geltung einer Regel bedeuten soll, zunächst offen bleiben kann).

[1] In der philosophischen Literatur hat insbesondere L. Wittgenstein dieses Verständnis von Regeln hervorgehoben; eine Einführung findet man bei A. Kemmerling (1975).

[2] Wir setzen jedoch nicht voraus, dass es *unbestimmt* viele Anwendungsfälle gibt, und betrachten z.B. auch die therapeutische Empfehlung eines Arztes, während eines bestimmten Zeitraums regelmäßig bestimmte Medikamente einzunehmen, als eine Regel.

Wir verwenden also einen weit gefassten Regelbegriff, der sich nicht von vornherein nur auf einen bestimmten Typ von Regeln bezieht.

2. Unterscheidungen durch Verwendungszwecke. Man kann Regeln unter verschiedenen Aspekten unterscheiden. Geht man davon aus, dass Regeln Hilfsmittel zur theoretischen und praktischen Orientierung sind, liegt es nahe, auf unterschiedliche Verwendungszwecke zu achten. Ohne Anspruch auf Vollständigkeit können in dieser Hinsicht folgende Unterscheidungen getroffen werden:

– *Normative Regeln*, durch die festgelegt wird, wie sich Menschen oder andere Lebewesen oder Dinge verhalten sollen oder wie sie beschaffen sein sollen.

– *Konstitutive Regeln*, durch die Bedeutungen von Objekten, Verhaltensweisen oder Situationen festgelegt werden.

– *Prognostische Regeln*, durch die angegeben wird, wie sich bestimmte, in der menschlichen Erfahrungswelt identifizierbare Objekte oder Sachverhalte (unter bestimmten Bedingungen) verhalten.

– *Poietische Regeln*, durch die angegeben wird, wie Menschen durch ihre Tätigkeiten bestimmte Sachverhalte bewirken können.

– *Nomologische Regeln*, die zur Konstruktion von (dynamischen) Modellen verwendet werden.

Im Folgenden werden wir einige dieser unterschiedlichen Arten von Regeln etwas näher besprechen. Dabei wird sich auch zeigen, dass sich die angeführten Charakterisierungen nicht unbedingt ausschließen. Nomologische Regeln werden erst später bei der Diskussion von Modellen besprochen.

3. Normative Regeln. Allgemein kann man von einer *normativen Aussage* (oder gleichbedeutend von einer *Norm*) sprechen, wenn durch die Aussage angegeben wird, wie ein Objekt, ein Sachverhalt, ein Verhalten oder eine Tätigkeit, ggf. unter bestimmten Bedingungen, beschaffen sein soll. Um eine normative Regel handelt es sich, wenn sich die normative Aussage auf (unbestimmt) viele Anwendungsfälle bezieht. Offenbar gibt es auch normative Aussagen, die keine normativen Regeln sind, wie z.B. die normativen Aussagen in einem Pflichtenheft, das sich auf die Herstellung eines singulären Artefakts bezieht. Uns interessieren hier nur normative Regeln. Exemplarisch kann man an Verkehrsregeln denken, z.B. an die Regel, dass Autofahrer vor einer roten Ampel anhalten sollen. Natürlich können normative Regeln mehr oder weniger große Spielräume für situationsabhängige Entscheidungen vorsehen, wie man sich z.B. anhand von Arbeitsverträgen verdeutlichen kann.

Es muss auch bedacht werden, dass sich normative Regeln nicht nur auf menschliches Verhalten beziehen können. So gibt es z.B. normative Regeln, die sich auf das Verhalten oder auf Eigenschaften von Tieren beziehen. Weiterhin kann man an technische Normen denken, durch die Eigenschaften oder Verhaltensweisen technischer Artefakte festgelegt werden. Zwar bedarf es stets menschlicher Tätigkeiten, um das durch die normative Regel geforderte Verhalten zu realisieren (z.B. durch eine der Norm entsprechende Konstruktion und Verwendung der Artefakte); insofern sind die *Subjekte normativer Regeln* (wie übrigens auch aller anderen Arten von Regeln) stets Menschen. Davon unabhängig kann jedoch von den Sachverhalten oder Verhaltensweisen gesprochen werden, die durch eine Regel normiert werden sollen; sie werden im Folgenden als *thematischer Bezug einer normativen Regel* bezeichnet.[3]

Von grundsätzlicher Bedeutung ist, dass durch normative Regeln keine Aussagen über die Beschaffenheit von Sachverhalten in der menschlichen Erfahrungswelt getroffen werden. Eine normative Regel legt fest, wie ein Sachverhalt oder ein Verhalten (unter bestimmten Bedingungen) beschaffen sein *soll*. Es ist auch klar, dass das, was durch eine normative Regel gefordert wird, nicht durch die Regel selbst bewirkt werden kann. Zum Beispiel kann die normative Regel, dass Autofahrer vor einer roten Ampel anhalten sollen, nicht bewirken, dass dies auch tatsächlich geschieht.

4. Unterschiedliches Reden von Normen. In der sozialwissenschaftlichen Literatur wird in unterschiedlichen Bedeutungen von „(sozialen) Normen“ gesprochen.

a) In einer ersten Bedeutungsvariante sind normative Regeln gemeint, allerdings oft mit einem auf menschliches Verhalten eingeschränkten thematischen Bezug. Dem entspricht z.B. folgende Erläuterung von J. Blake und K. Davis in einem Handbuch-Artikel über „Norms, Values, and Sanctions“ (1964: 456): „the term [norm] is employed [...] to designate any standard or rule that states what human beings should or should not think, say, or do under given circumstances.“ Ähnliche Definitionen findet man auch bei anderen Autoren.[4]

b) Eine zweite Bedeutungsvariante knüpft an empirische Aussagen über

[3] Auf diesen thematischen Bezug bezieht sich wohl auch meistens die Unterscheidung zwischen *sozialen* Regeln (für das Verhalten von Menschen) und *technischen* Regeln (für das Verhalten und die Eigenschaften technischer Artefakte); es gibt jedoch auch andere Unterscheidungsvorschläge, z.B. bei B. Joerges (1989).

[4] Außer der Einschränkung des thematischen Bezugs auf menschliches Verhalten wird von einigen Autoren auch vorausgesetzt, dass Normen „gelten“. So heißt es z.B. bei H. P. Bahrdt (1994: 49): „Normen sind allgemein geltende und in ihrer Allgemeinheit verständlich mitteilbare Vorschriften für menschliches Handeln [...].“ Die Annahme, dass Normen „gelten“, definitorisch vorauszusetzen, ist jedoch unzweckmäßig, denn sie verstellt die Frage, ob und ggf. wie Normen „gelten“.

normative Meinungen an, die sich einzelnen Menschen zurechnen lassen. Als Beispiel sei auf K.-D. Opp hingewiesen, der unter Normen „geäußerte Erwartungen“ versteht (1983: 4). Tatsächlich meint Opp nicht Erwartungen in der üblichen Bedeutung dieses Worts, sondern normative Meinungen,[5] wie folgende Erläuterung zeigt:

> „Zur Vermeidung von Mißverständnissen sei nachgetragen, daß Erwartungen *in verschiedenen Situationen* geäußert werden können. Wenn z.B. eine Person meint, man solle sonntags in die Kirche gehen, dann mag sie dies einem Interviewer sagen, sie mag dies gegenüber jemandem, der kein „Kirchgänger“ ist, zum Ausdruck bringen, oder auch in einem Gespräch über religiöse Fragen äußern.“ (Opp 1983: 15)

Bei diesem Ansatz werden also Aussagen über Normen als Aussagen über normative Meinungen oder Überzeugungen, die man bei den Mitgliedern einer Gesellschaft ermitteln kann, aufgefasst.

c) Eine dritte Variante knüpft an Vorstellungen über Verhaltensregelmäßigkeiten an. Folgende Formulierung von H. Popitz (1967: 22) verdeutlicht diesen Sprachgebrauch:

> „Als Verhaltensnormen bezeichnen wir Verhaltensweisen, die von allen oder einer bestimmten Kategorie von Gesellschafts- bzw. Gruppenmitgliedern in einer bestimmten Situation regelmäßig wiederholt und im Fall der Abweichung durch eine negative Sanktion gegen den Abweicher bekräftigt werden. Wir beziehen uns also auf ein *tatsächlich ablaufendes Verhalten,* nicht auf ein gewünschtes oder als verbindlich gedachtes und auch nicht auf subjektiv erwartetes Verhalten.“

Es gibt also mindestens drei unterschiedliche Weisen, in denen von (sozialen) Normen gesprochen wird. Deshalb sei noch einmal betont, dass in diesem Text das Wort ‘Norm’ gleichbedeutend mit ‘normative Aussage’ verwendet wird. Somit können auch normative Regeln als Normen, die in der sprachlichen Form einer Regel formuliert sind, aufgefasst werden.

5. Konstitutive Regeln. Die Idee, konstitutive Regeln als einen eigenständigen Typ von Regeln zu konzipieren, wurde insbesondere von John R. Searle verbreitet. In einer frühen Arbeit über Sprechakte hat Searle den Grundgedanken folgendermaßen erläutert:

> „I want to clarify a distinction between two different sorts of rules, which I shall call *regulative* and *constitutive* rules. I am fairly confident about the distinction, but do not find it easy to clarify. As a start, we might say that regulative rules regulate antecedently or independently existing forms of behavior; for example, many rules of etiquette regulate inter-personal relationships which exist independently of the rules. But constitutive rules do not merely regulate, they create or define new forms of behavior. The rules of football or chess, for example, do

[5] Das potentiell irreführende Konfundieren von normativen Meinungen und „Erwartungen“ findet man in der soziologischen Literatur häufiger; z.B. beginnt F. M. Cancian (1975: 1) ihre Untersuchung über Normen mit der Bemerkung: „Norms can be loosely defined as shared conceptions of appropriate or expected action.“

not merely regulate playing football or chess, but as it were they create the very possibility of playing such games. The activities of playing football or chess are constituted by acting in accordance with (at least a large subset of) the appropriate rules. Regulative rules regulate a pre-existing activity, an activity whose existence is logically independent of the rules. Constitutive rules constitute (and also regulate) an activity the existence of which is logically dependent on the rules.“ (Searle 1969: 33f.)

Allerdings gelangt man nicht ohne weiteres zu einer klaren Unterscheidung. Zum Beispiel gibt es Regeln für Fussballspiele; aber wäre es unmöglich, Fußball zu spielen, wenn es diese Regeln nicht gäbe? Zwar ist richtig, dass man ohne diese Regeln nicht *ihnen gemäß* Fußball spielen könnte, aber dieses Argument könnte ebenso bei regulativen Regeln angeführt werden. Gäbe es beispielsweise keine Verkehrsregeln, könnte man sich im Straßenverkehr nicht *ihnen gemäß* verhalten. Insofern liefert Searles Überlegung keine ausreichende Begründung für eine Unterscheidung zwischen regulativen und konstitutiven Regeln.[6]

Um zu einer sinnvollen Definition zu gelangen, kann jedoch an folgende sprachliche Form angeknüpft werden, die Searle (1969: 35) für konstitutive Regeln vorgeschlagen hat: X counts as Y in context C. Orientiert man sich an diesem Schema, dienen konstitutive Regeln dem Zweck, Objekten, Verhaltensweisen oder Situationen Bedeutungen zu verleihen. Zum Beispiel: Definitionen sprachlicher Ausdrücke, die Regeln des ASCII-Codes zur Fixierung der Bedeutung von Bit-Mustern und die Regeln, durch die die Bedeutung von Lichtsignalen im Straßenverkehr oder von Schallsignalen in der Schiffahrt festgelegt werden. Versteht man konstitutive Regeln in diesem Sinn, kann man auch noch einmal die anfangs zitierte Überlegung von Searle verfolgen. Denn man kann offenbar sagen, dass es durch konstitutive Regeln möglich wird, Tätigkeiten, Objekten oder Situationen eine intersubjektiv vermittelbare Bedeutung *zu geben*.[7] Wir werden im Weiteren von dieser Definition ausgehen; wenn von konstitutiven Regeln gesprochen wird, sind also stets Regeln gemeint, durch die Bedeutungen festgelegt werden.

6. Prognostische und poietische Regeln. Im Unterschied zu normativen und konstitutiven Regeln dienen prognostische Regeln dem Zweck, einschätzbar zu machen, was unter bestimmten Bedingungen wahrscheinlich der Fall gewesen ist oder sein wird. Prognostische Regeln beziehen sich also auf Prozesse in der menschlichen Erfahrungswelt. Instruktiv ist ein Vergleich mit normativen Regeln, zum Beispiel:

> (1) Ein Autofahrer *soll* vor einer roten Ampel anhalten. (2) Ein Autofahrer *wird (wahrscheinlich)* vor einer roten Ampel anhalten.

[6] Man vgl. hierzu auch die Kritik bei M. Black (1962: 123f.).

[7] Allerdings können die von Searle angeführten Beispiele (Spielregeln für Schach und Fußball) nicht mehr ohne weiteres als konstitutive Regeln verstanden werden.

Im ersten Fall handelt es sich um eine normative, im zweiten Fall um eine prognostische Regel. Es ist offensichtlich, dass weder (2) aus (1) noch umgekehrt (1) aus (2) ableitbar ist.

Eine wichtige Variante prognostischer Regeln entsteht, wenn das Zustandekommen des durch die Regel vorausgesagten Sachverhalts auch von Tätigkeiten derjenigen Person (oder Personen) abhängt, auf die als Subjekt der Regelverwendung Bezug genommen wird. Solche Regeln zeigen, wie Menschen durch Tätigkeiten bestimmte Wirkungen erzielen können, und ich nenne sie deshalb *poietische Regeln*; zum Beispiel: Indem man ein Streichholz an einer rauhen Fläche reibt, kann man es zur Entzündung bringen. Offenbar hängt das Ergebnis auch davon ab, wie sich ein menschlicher Akteur verhält.

So allgemein formuliert gilt die Aussage allerdings auch für viele prognostische Regeln, die keine poietischen Regeln sind, z.B. für die oben unter (2) angegebene Regel. Denn was geschehen wird, wenn sich ein Autofahrer einer roten Ampel nähert, hängt offenbar auch und vor allem von seinem Verhalten ab. Aber im Unterschied zum Streichholz-Beispiel kann man mit dieser Regel keine Wirkungen hervorbringen; das Geschehen, auf das sich die Regel bezieht, läuft im einen Fall unabhängig, im anderen Fall nicht unabhängig von dem Verhalten *derjenigen* Personen ab, die sich mithilfe der Regel orientieren. Also kann man unterscheiden:

- Prognostische Regeln, die aus einer Beobachterperspektive formuliert werden können; und
- poietische Regeln, die aus einer Akteursperspektive formuliert werden müssen und bei denen das Geschehen, auf das sich die Regel bezieht, auch von Tätigkeiten eines als Akteur konzipierten Subjekts der Regelverwendung abhängt.

Natürlich darf nicht vergessen werden, dass auch bei vielen (nicht allen) prognostischen Regeln, die aus einer Beobachterperspektive formuliert werden können, das Geschehen, auf das sich die Regel bezieht, von menschlichen Tätigkeiten abhängt.

7. Poietische und normative Regeln. Obwohl poietische Regeln als ein Sonderfall prognostischer Regeln betrachtet werden können, muss andererseits bedacht werden, dass sie auch Ähnlichkeiten mit normativen Regeln aufweisen. Man muss es „richtig machen“, um mithilfe und entsprechend der Regel die beabsichtigte Wirkung zu erzielen. Darauf zu achten, ist wichtig, denn das Kennen- und Verwendenlernen poietischer Regeln ist für das menschliche Leben von grundlegender Bedeutung. Man kann wohl vermuten, dass sich die meisten Menschen in den meisten ihrer Tätigkeiten an vorgängig gelernten poietischen Regeln orientieren. Bei dieser Aussage setze ich natürlich einen weit gefassten Begriff poietischer Regeln voraus und beziehe mich nicht nur auf technische Regeln. Zum Beispiel kann man auch an Regeln für Fussball- oder Schachspiele, an Regeln für korrektes

Sprechen und Schreiben und an Regeln zur Berechnung statistischer Mittelwerte und Korrelationen denken. All dies sind poietische Regeln, die *auch* als normative Regeln betrachtet werden können.

Insofern wäre es auch falsch, poietische und normative Regeln als unterschiedliche Arten von Regeln zu unterscheiden. Denn die meisten normativen Regeln können auch als poietische Regeln betrachtet werden. Die Bedeutung dieser Betrachtungsweise wird sichtbar, wenn man darauf achtet, dass den Geltungsansprüchen normativer Regeln keine kausale Wirksamkeit zukommt, sondern dass ihre Realisierung voraussetzt, dass menschliche Akteure ihr Verhalten an den normativen Regeln orientieren.

8. Regeln kennen, verwenden, beachten, befolgen. Regeln können für unterschiedliche Zwecke verwendet werden. Allgemein können sie als gedankliche Hilfsmittel zur Orientierung im Denken und Handeln charakterisiert werden. Je nachdem, um welche Arten von Regeln es sich handelt, können die Verwendungsweisen auch spezifischer charakterisiert werden. So kann man bei prognostischen Regeln davon sprechen, dass sie für Voraussagen verwendet werden können; andererseits kann man bei normativen Regeln oft davon sprechen, dass sich Menschen ihnen entsprechend oder von ihnen abweichend verhalten können. Die Formulierung, dass Menschen in ihrem Verhalten Regeln „folgen", trifft also keineswegs alle Formen, in denen Menschen Regeln unterschiedlicher Art verwenden.

Prognostische Regeln können auch zur Bildung von Erwartungen verwendet werden.[8] Hierin liegt einer der Unterschiede zu normativen Regeln, die nicht ohne weiteres zur Bildung von Erwartungen verwendet werden können. Zum Beispiel liefert die normative Regel, dass Autofahrer vor einer roten Ampel anhalten sollen, für sich genommen keine Informationen darüber, wie sich Autofahrer vor roten Ampeln tatsächlich verhalten. Man benötigt vielmehr eine korrespondierende prognostische Regel, die sich auf das tatsächliche Verhalten von Autofahrern bezieht.

Wie auch immer Regeln verwendet werden, ihre Verwendung setzt ihre Kenntnis voraus. Damit ein Mensch eine Regel zur Orientierung im Denken und Handeln verwenden kann, muss er sie kennen. Damit ist zwar nicht gemeint, dass Regeln stets bewusst verwendet werden müssen; denn es gibt zahlreiche Beispiele, in denen die Orientierung an einer Regel – nachdem sie gelernt worden ist – in Form einer Gewohnheit, ohne ein explizites Bewusstsein der Regelanwendung geschieht. Dies setzt aber nicht nur ein vorgängiges Erlernen der Regel voraus, sondern auch, dass die Regel bei Bedarf formuliert und zum Gegenstand von Überlegungen gemacht werden kann. Mindestens in diesem Sinn muss ein Mensch eine Regel kennen, um

[8] Wir unterscheiden zwischen Voraussagen und Erwartungen. *Voraussagen* sind Aussagen und haben als solche keinen bestimmten individuellen Subjektbezug. Dagegen beziehen sich *Erwartungen* darauf, was jeweils bestimmte Menschen in bestimmten Situationen erwarten. Während Voraussagen (post factum) wahr oder falsch sein können, werden Erwartungen erfüllt oder enttäuscht.

davon sprechen zu können, dass er die Regel verwenden, beachten, befolgen oder sich an ihr orientieren kann.

Es soll also nicht genügen, dass ein Mensch sich einer Regel entsprechend verhält. Das kann ggf. aus einer Beobachterperspektive festgestellt werden, liefert aber für sich genommen noch keinen hinreichenden Grund für die Annahme, dass der betreffende Mensch sein Verhalten an einer Regel orientiert. Dem entspricht eine Unterscheidung zwischen „fitting" und „guiding" bei W. V. Quine (1972: 442): „Behavior *fits* a rule whenever it conforms to it; whenever the rule truly decribes the behavior. But the behavior ist not *guided* by the rule unless the behaver knows the rule and can state it. This behaver *observes* the rule." Ähnliche Unterscheidungen wurden von anderen Autoren vorgeschlagen, u.a. von R. D. Gumb (1972: 41).[9]

4.2 Nomologische dynamische Modelle

1. Modelle und modale Fragestellungen. In der Literatur findet man oft die Auffassung, dass Modelle als „vereinfachende Beschreibungen" von Ausschnitten der menschlichen Erfahrungswelt verstanden werden können; hier sind zwei typische Formulierungen:

> „A *scientific model* is an abstract and simplified description of a given phenomena." (Olkin, Gleser und Derman 1980: 2) „A model of any set of phenomena is a formal representation thereof in which certain features are abstracted while others are ignored with the intent of providing a simpler description of the salient aspects of the chosen phenomena." (Hendry und Richard 1982: 4)

Sehr ähnlich sind Formulierungen, in denen von „Abbildungen" gesprochen wird, zum Beispiel:

> „Modelle können wir uns in erster Näherung denken als begriffliche Konstrukte zur 'Abbildung' realer Systeme oder zum Umgang mit solchen." (Balzer 1997: 16) „Ein Modell ist wohl immer aufzufassen als eine Abbildung. Die Frage ist nur, was abgebildet wird, und wie die Abbildungsfunktion aussieht." (Frey 1961: 89) „Ein Modell ist eine Abbildung von für die jeweilige Fragestellung bedeutsamen Teilaspekten der Wirklichkeit zu einem vereinfachten System." (Wirth 1979: 130f.) „Ein Modell ist immer eine vereinfachte Abbildung eines interessierenden Realitätsausschnitts." (Bossel 1992: 27)

In vielen, vermutlich sogar den meisten Fällen haben jedoch Modelle, wie sie in der sozialwissenschaftlichen Literatur konstruiert und diskutiert werden, nicht die Aufgabe, Ausschnitte der menschlichen Erfahrungswelt zu beschreiben oder „abzubilden". Beispielsweise kann man an demographische Modelle denken, die dem Zweck dienen, *mögliche* Bevölkerungsentwicklungen vorstellbar zu machen, die aus hypothetischen Annahmen über

[9] Dagegen ist eingewendet worden, dass Menschen ihr Verhalten an einer Regel orientieren (und insofern die Regel „kennen") können, ohne unbedingt in der Lage zu sein, eine explizite sprachliche Formulierung der Regel anzugeben. Eine ausführliche Diskussion findet man bei John Fisher (1975).

Geburten, Todesfälle und Migrationen ableitbar sind. Offenbar kann man nicht sagen, dass durch Modelle dieser Art reale historische Prozesse beschrieben werden. Dieses Beispiel liefert auch einen allgemeinen Gesichtspunkt: In vielen Fällen dienen Modelle dem Nachdenken über Möglichkeiten; oder in einer kurzen Formulierung: *Modelle sind Hilfsmittel zur Reflexion modaler Fragestellungen,* wobei sich diese Fragestellungen sowohl auf zukünftige Möglichkeiten als auch auf die Beschaffenheit bereits realisierter Sachverhalte, über die nur unzureichende Informationen verfügbar sind, beziehen können.[10] Die als Leitfaden dienenden modalen Fragestellungen können natürlich sehr unterschiedlich sein und müssen bei der Konstruktion von Modellen erläutert werden.

2. Bezugsprobleme dynamischer Modelle. Wenn im Folgenden von *dynamischen Modellen* gesprochen wird, sind Modelle gemeint, die dem Zweck dienen, (reale oder fiktive) Prozesse darzustellen und in ihrem Ablauf einschätzbar zu machen. Etwas genauer können drei Zwecke unterschieden werden:

- Der Zweck eines dynamischen Modells kann in erster Linie darin bestehen, mit seiner Hilfe bestimmte Vorstellungen über einen fiktiven oder in der Vergangenheit tatsächlich abgelaufenen Prozess zu gewinnen.
- Ein dynamisches Modell kann dem Zweck dienen, hypothetische Aussagen über den Ablauf eines (oft, aber nicht unbedingt zukünftigen) Prozesses zu formulieren und deutlich zu machen, wie sie von Bedingungen abhängig sind.
- Ein dynamisches Modell kann hauptsächlich dazu dienen, einen analytischen Rahmen bereitzustellen, um darüber nachzudenken, wie und wodurch mögliche Prozesse zustandekommen und von Bedingungen abhängen.

Offenbar handelt es sich um mögliche Zwecke dynamischer Modelle, die sich nicht ausschließen.

3. Ein formaler Rahmen für dynamische Modelle. Die Konstruktion dynamischer Modelle hängt in erster Linie von der Konzeptualisierung der zu modellierenden Prozesse ab. Einen sehr allgemeinen Modellrahmen liefert das in Abschnitt 3.2 besprochene Zeitreihenschema. Ausgangspunkt ist eine diskrete Zeitachse $\mathcal{T}$, die als eine geordnete Menge von Zeitstellen konzipiert und formal mit der Menge der ganzen Zahlen identifiziert wird. Zur Definition von Zeitreihen wird eine zusammenhängende Teilmenge $\mathcal{T}^* \subset \mathcal{T}$ verwendet, wobei die Anzahl der Zeitstellen in $\mathcal{T}^*$ endlich oder unendlich sein kann. In jedem Fall gibt es jedoch eine kleinste Zeitstelle, so dass man auch eine Prozesszeitachse $\mathcal{T}^* := \{0, 1, 2, 3, \ldots\}$ verwenden kann. Hiervon ausgehend kann das allgemeine Zeitreihenschema durch eine Funktion $X : \mathcal{T}^* \longrightarrow \tilde{\mathcal{X}}$ dargestellt werden. Eine Zeitreihe X ist somit formal eine Funktion, die jeder Zeitstelle $t \in \mathcal{T}^*$ einen Wert $X(t)$ in einem Wertebereich $\tilde{\mathcal{X}}$ zuordnet. Dabei gibt es keinerlei einschränkende Bedingungen für die Funktion X bzw. ihren Wertebereich $\tilde{\mathcal{X}}$. Der Wertebereich kann mehrdimensional sein, und es ist auch möglich, zeitliche Folgen statistischer Variablen zu erfassen (wobei dann die Elemente von $\tilde{\mathcal{X}}$ ebenfalls Funktionen sind).

[10] In ähnlicher Weise hat J. Heckman (2005: 2-3) wissenschaftliche Modelle charakterisiert.

Durch das Zeitreihenschema erhält man also zunächst einen allgemeinen begrifflichen Rahmen zur Konzeption möglicher Prozesse.

4. Nomologische und poietische Modelle. Um zu dynamischen Modellen zu gelangen, sind Annahmen über das Zustandekommen der zu modellierenden Prozesse erforderlich. Dabei kann man von zwei unterschiedlichen Ideen ausgehen:

- Man kann von der Idee ausgehen, dass Prozessabläufe „durch Regeln bestimmt“ werden; wir sprechen dann von *nomologischen Modellen*.
- Man kann von der Idee ausgehen, dass Prozesse durch Tätigkeiten von Akteuren zustandekommen; dementsprechend kann man von *poietischen Modellen* sprechen.

Wie bereits in Abschnitt 3.1 (§ 5) bemerkt wurde, ist die Formulierung, dass Prozesse teilweise oder vollständig „durch Regeln bestimmt“ sind, ambivalent; es sei deshalb noch einmal betont: Wenn bei einem nomologischen Modell von „durch Regeln bestimmten“ Prozessen gesprochen wird, bezieht sich diese Aussage auf die mithilfe des Modells konstruierten Prozesse, also auf die Prozesse einer artifiziellen Modellwelt (in der, wie man auch sagen könnte, die Regeln gleichsam als Gesetze gelten, die der Modellkonstruktur seiner Modellwelt vorgeschrieben hat). Aussagen dieser Art implizieren infolgedessen nicht, dass auch die in der Realität ablaufenden Prozesse „durch Regeln bestimmt“ sind.

5. Deterministische und stochastische Modelle. Bei nomologischen Modellansätzen geht man von der Vorstellung aus, dass die zu modellierenden Prozesse nach gewissen Regeln ablaufen. Zur Konzeption solcher Regeln gibt es hauptsächlich zwei unterschiedliche Möglichkeiten.

- *Deterministische Regeln* liefern für jeden möglichen Zustand auf eindeutige Weise einen nachfolgenden Zustand.
- *Stochastische Regeln* liefern für jeden möglichen Zustand eine Wahrscheinlichkeitsverteilung für mindestens zwei mögliche nachfolgende Zustände.

Dementsprechend können *deterministische* und *stochastische Modelle* unterschieden werden. Wir betrachten dies als zwei Varianten nomologischer Modelle, identifizieren also nomologische Modelle nicht von vornherein mit deterministischen Modellen.

6. Offene und geschlossene Modellkonzeptionen. Eine wichtige Unterscheidung bezieht sich darauf, ob bzw. wie durch ein Modell Kontexte der Prozessabläufe berücksichtigt werden. Ein *geschlossenes dynamisches Modell* kann allgemein durch ein Schema der folgenden Art dargestellt werden:

$$X(0) \longrightarrow X(1) \longrightarrow X(2) \longrightarrow X(3) \longrightarrow X(4) \longrightarrow$$

Dabei soll der Pfeil $\longrightarrow$ einen zeitlichen Übergang andeuten, der „durch eine Regel bestimmt“ wird. Von einem *geschlossenen* Modell wird gesprochen, weil nur der Anfangszustand $X(0)$ nicht innerhalb des Modells bestimmt wird; oder anders formuliert: Mit Ausnahme des Anfangszustands hängt der Prozess nur von Bedingungen ab, deren Zustandekommen innerhalb des Modells erklärt werden kann.

Ein *offenes dynamisches Modell* kann dagegen in folgender Form dargestellt werden:

$$\begin{tikzcd}
X(0) \arrow[r] & X(1) \arrow[r] & X(2) \arrow[r] & X(3) \arrow[r] & X(4) \arrow[r] & {} \\
Y(0) \arrow[ur] & Y(1) \arrow[ur] & Y(2) \arrow[ur] & Y(3) \arrow[ur] & &
\end{tikzcd}$$

In diesem Fall hängt der durch $X(t)$ repräsentierte Prozess auch von den Zuständen $Y(t)$ eines anderen (Teil-) Prozesses ab, über deren Zustandekommen innerhalb des Modells keine Annahmen getroffen werden.[11] Um zu bestimmten Verläufen des X-Prozesses zu gelangen, muss also in diesem Fall nicht nur der Anfangszustand $X(0)$, sondern außerdem der gesamte Y-Prozess vorausgesetzt werden.[12]

Eine weitere Differenzierungsmöglichkeit betrifft die Zeitabhängigkeit der Regeln. Bei den bisher verwendeten schematischen Formulierungen für geschlossene Modelle hängt $X(t+1)$ nur vom vorangehenden Zustand $X(t)$ ab, frühere Zustände sind jedenfalls nicht direkt relevant. Offenbar kann man auch annehmen, dass $X(t+1)$ nicht nur von $X(t)$, sondern auch von $X(t-1), \ldots, X(t-k)$ abhängt. Wir sprechen dann von einer *Regel mit einem k-stufigen Gedächtnis*; als Spezialfall erhält man eine *gedächtnislose*

[11] Offenbar kann man wahlweise von zwei separaten Prozessen oder von zwei Teilprozesses eines gemeinsamen Prozesses sprechen.

[12] Es sei betont, dass sich hier die Adjektive ‘offen’ und ‘geschlossen’ auf Modelle beziehen, nicht jedoch auf „reale Systeme“, wie z.B. in der folgenden Überlegung von W. Stegmüller (1983: 156): „Ein System kann *absolut abgeschlossen* sein, wenn die darin vorkommenden Prozesse überhaupt keiner äußeren Einwirkung unterliegen. Strenggenommen gibt es nur ein einziges derartiges System: das Universum. Ein System S ist *relativ abgeschlossen*, wenn entweder die äußeren Einwirkungen für die Prozesse innerhalb von S ohne Relevanz sind oder wenn sie konstanter Natur sind, so dass man sie bei der Formulierung von Gesetzen für die Übergänge von Zuständen von S vernachlässigen kann.“ Somit liegt die von Stegmüller nicht diskutierte Frage nahe, ob es wenigstens „relativ abgeschlossene“ reale Systeme gibt. Sie kann jedoch vermieden werden, wenn man sich von vornherein nur auf Modelle bezieht.

Regel, wenn $k = 0$ ist.[13] (Analoge Definitionen können für offene Modelle verwendet werden.)

4.3 Stochastische Ablaufschemas

In diesem Abschnitt werden einfache stochastische Ablaufschemas definiert und wird die Grundidee stochastischer Modelle besprochen, die an solche Ablaufschemas anknüpfen. Als Hilfsmittel dienen Zufallsgeneratoren und Zufallsvariablen, deren Begriffe zunächst erläutert werden.

1. Zufallsgeneratoren. Von Wahrscheinlichkeiten wird in unterschiedlichen Bedeutungen gesprochen; im Folgenden sind stets aleatorische Wahrscheinlichkeiten gemeint, die sich durch einen gedanklichen Rückgriff auf *Zufallsgeneratoren* definieren lassen. Folgendes Bild kann zur Erläuterung dienen.[14]

$$\text{Aktivierung} \longrightarrow \boxed{\text{Zufallsgenerator}} \longrightarrow \text{Sachverhalt}$$

a) Ein Zufallsgenerator ist ein *Verfahren*, um Sachverhalte zu erzeugen. Das impliziert, dass ein Zufallsgenerator nicht selbst ein Akteur ist, vielmehr einen Akteur voraussetzt, der das Verfahren anwendet, um einen Sachverhalt zu erzeugen. In dem Bild wird dies durch das Wort ‘Aktivierung’ angedeutet. Exemplarisch kann man daran denken, dass zur Definition des Zufallsgenerators ein Würfel verwendet wird: Irgendjemand muss ihn nehmen und werfen, damit ein neuer Sachverhalt entsteht.

b) Die Beschreibung eines Zufallsgenerators besteht infolgedessen in der Beschreibung eines Verfahrens zur Erzeugung von Sachverhalten. Dazu gehört auch eine Beschreibung von ggf. zu verwendenden Geräten, etwa eines Würfels, wenn ein solcher verwendet werden soll. Wichtig ist, dass sich der Begriff im Wesentlichen auf Regeln bezieht, die ein Verfahren charakterisieren. Soll z.B. ein Würfel verwendet werden, genügt es nicht, den Würfel zu definieren, sondern es muss außerdem

[13] Auf den ersten Blick scheinen sich die Möglichkeiten zur Formulierung nomologischer Modelle durch die Verwendung von Regeln mit einem Gedächtnis wesentlich zu vergrößern. Es ist deshalb bemerkenswert, dass das in gewisser Weise nicht der Fall ist. Genauer gesagt: Ein nomologisches Modell, in dem Regeln mit einem k-stufigen Gedächtnis verwendet werden, kann stets in ein Modell mit gedächtnislosen Regeln transformiert werden. Es ist nur erforderlich, den Zustandsraum $\tilde{\mathcal{X}}$ durch ein Gedächtnis (formal identisch mit $\tilde{\mathcal{X}}^k$) zu ergänzen; dann kann offenbar die Abhängigkeit von den vergangenen k Zuständen auch als eine Abhängigkeit vom gegenwärtigen Zustand des Gedächtnisses dargestellt werden.

[14] Wir folgen hier den Ausführungen bei Rohwer und Pötter (2001: 264ff.). Eine Diskussion unterschiedlicher Wahrscheinlichkeitsbegriffe findet man bei Rohwer und Pötter (2002b).

angegeben werden, *wie* der Würfel verwendet werden soll. Der Begriff 'Zufallsgenerator' meint in diesem Fall nicht nur den Würfel, sondern auch die Art und Weise, wie mit dem Würfel Sachverhalte zu erzeugen sind.

c) Dass es sich bei einem Zufallsgenerator um ein Verfahren handelt, impliziert weiterhin, dass man es wiederholt (im Prinzip beliebig oft) verwenden kann, um immer neue Sachverhalte zu erzeugen. Man kann z.B. einen Würfel beliebig oft verwenden, d.h. ihn werfen und dadurch einen jeweils neuen Sachverhalt erzeugen.

d) Durch die Aktivierung eines Zufallsgenerators können Sachverhalte unterschiedlichen Typs entstehen. Zum Beispiel kann man bei der Verwendung eines Würfels sechs unterschiedliche Typen von Sachverhalten festlegen, entsprechend den sechs möglichen Augenzahlen. Die Formulierung, dass Sachverhalte unterschiedlichen Typs „entstehen können", soll bedeuten: zum Zeitpunkt der Aktivierung eines Zufallsgenerators ist der Typ des daraus resultierenden Sachverhalts noch unbestimmt.

e) Der Prozess, der bei der Aktivierung eines Zufallsgenerators zu einem jeweils bestimmten Sachverhalt führt, soll unabhängig davon verlaufen, wann, wo und von wem der Zufallsgenerator aktiviert wird. Dies kann natürlich nur als eine Forderung an einen idealen Zufallsgenerator verstanden werden. Als Forderung ist sie jedoch wesentlich, denn sie rechtfertigt es, dass man bei der Charakterisierung eines Zufallsgenerators auf alle Arten von Bezugnahmen auf einen historischen Prozess, in dessen Rahmen der Zufallsgenerator verwendet wird, absehen kann. Insbesondere impliziert die Forderung, dass der Akteur, der den Zufallsgenerator aktiviert, keinen Einfluss darauf nehmen kann, welcher Sachverhalt entstehen wird.[15]

f) Eine weitere, in gewisser Weise bereits implizierte Forderung besteht schließlich darin, dass der Prozess, der bei der Aktivierung eines Zufallsgenerators zu einem jeweils bestimmten Sachverhalt führt, auch unabhängig davon verläuft, welche Sachverhalte zuvor mit ihm erzeugt worden sind. Wiederum handelt es sich um eine Forderung an einen idealen Zufallsgenerator: er soll kein Gedächtnis haben; oder anders formuliert: die Funktionsweise eines Zufallsgenerators soll hinreichend charakterisierbar sein, ohne dabei auf die Geschichte seiner bisherigen Verwendung Bezug nehmen zu müssen.

2. Aleatorische Wahrscheinlichkeit. Die Begriffsbildungen der Wahrscheinlichkeitsrechnung knüpfen an die eben genannten Eigenschaften eines Zufallsgenerators an. In einem ersten Schritt wird festgelegt, welche Arten von Sachverhalten als möglich betrachtet werden sollen. Zu diesem

[15] Orientiert man sich an der in Abschnitt 3.1 (§ 4, Anm. 9) gegebenen Erläuterung, löst die Aktivierung eines Zufallsgenerators einen mechanischen Prozess aus.

Zweck wird ein *Merkmalsraum* $\tilde{\mathcal{Z}}$ fixiert, dessen Elemente verwendet werden können, um die Sachverhalte zu charakterisieren, die bei der Aktivierung eines Zufallsgenerators entstehen können. Wird z.B. für den Zufallsgenerator ein Würfel verwendet, kann man den Merkmalsraum $\tilde{\mathcal{Z}} := \{\tilde{z}_1, \tilde{z}_2, \tilde{z}_3, \tilde{z}_4, \tilde{z}_5, \tilde{z}_6\}$ verwenden, wobei $\tilde{z}_j$ bedeuten soll, dass als Ergebnis eines Wurfs die Augenzahl j erscheint. Die Notation $\tilde{\mathcal{Z}}$ soll darauf hinweisen, dass es eine Parallele zu den Merkmalsräumen statistischer Variablen gibt. Wie bei deren Merkmalsräumen wird auch bei den Merkmalsräumen für Zufallsgeneratoren davon ausgegangen, dass man sich stets eine numerische Repräsentation verschaffen kann. In unserem Beispiel könnte man den Merkmalsraum durch $\tilde{\mathcal{Z}} := \{1, 2, 3, 4, 5, 6\}$ fixieren und ergänzend vereinbaren, was die verwendeten Zahlen bedeuten sollen. Weiterhin kann man, ebenfalls analog zu den Merkmalsräumen statistischer Variablen, Teilmengen eines Merkmalsraums (Merkmalsmengen) verwenden, um Sachverhalte, die durch die Aktivierung eines Zufallsgenerators entstehen, zu charakterisieren. Zum Beispiel kann die Teilmenge $\tilde{Z} := \{2, 4, 6\} \subset \tilde{\mathcal{Z}}$ verwendet werden, um auszudrücken, dass eine gerade Augenzahl erschienen ist.

Zweitens wird angenommen, dass sich bei einem Zufallsgenerator jeder Teilmenge $\tilde{Z} \subseteq \tilde{\mathcal{Z}}$ eine (aleatorische) Wahrscheinlichkeit zuordnen lässt, mit der Sachverhalte entstehen können, die sich durch ein Element von $\tilde{Z}$ charakterisieren lassen. Dafür wird folgende Definition verwendet: Ein *Wahrscheinlichkeitsmaß* (auch *Wahrscheinlichkeitsverteilung* genannt) für einen Zufallsgenerator $\mathcal{G}$ mit einem Merkmalsraum $\tilde{\mathcal{Z}}$ ist eine Funktion

$$\Pr[\mathcal{G}] : \mathcal{P}(\tilde{\mathcal{Z}}) \longrightarrow \mathbf{R}$$

(Pr soll an 'Probability' erinnern), die jeder Teilmenge $\tilde{Z} \subseteq \tilde{\mathcal{Z}}$ eine Wahrscheinlichkeit $\Pr[\mathcal{G}](\tilde{Z})$ zuordnet und für die folgende Bedingungen gelten:

a) für alle $\tilde{Z} \subseteq \tilde{\mathcal{Z}}$: $0 \leq \Pr[\mathcal{G}](\tilde{Z}) \leq 1$

b) $\Pr[\mathcal{G}](\emptyset) = 0,\ \Pr[\mathcal{G}](\tilde{\mathcal{Z}}) = 1$

c) für alle $\tilde{Z}, \tilde{Z}' \subseteq \tilde{\mathcal{Z}}$: wenn $\tilde{Z} \cap \tilde{Z}' = \emptyset$, dann gilt:
$$\Pr[\mathcal{G}](\tilde{Z} \cup \tilde{Z}') = \Pr[\mathcal{G}](\tilde{Z}) + \Pr[\mathcal{G}](\tilde{Z}')$$

Ein Wahrscheinlichkeitsmaß $\Pr[\mathcal{G}]$ ist also in formaler Hinsicht durch die gleichen Regeln bestimmt wie eine Häufigkeitsfunktion. Insofern gibt es bei einer rein formalen Betrachtungsweise keinen Unterschied zwischen Häufigkeits- und Wahrscheinlichkeitsverteilungen. Allerdings ist die Bedeutung unterschiedlich und wir verwenden deshalb unterschiedliche Symbole: $P[X]$ für die Häufigkeitsfunktion einer statistischen Variablen X und $\Pr[\mathcal{G}]$ für die Wahrscheinlichkeitsverteilung eines Zufallsgenerators $\mathcal{G}$.

3. Zufallsvariablen. Um sich flexibel auf Zufallsgeneratoren beziehen zu können, dienen Zufallsvariablen. Allgemein verstehen wir unter einer *Zufallsvariablen* eine Funktion der Form

$$\dot{X} : \tilde{\mathcal{Z}} \longrightarrow \tilde{\mathcal{X}}$$

wobei $\tilde{\mathcal{Z}}$ der Merkmalsraum eines Zufallsgenerators und $\tilde{\mathcal{X}}$ irgendein Merkmalsraum ist, der auch mit $\tilde{\mathcal{Z}}$ identisch sein kann.

Ist die Zufallsvariable $\dot{X}$ durch eine Bezugnahme auf einen Zufallsgenerator $\mathcal{G}$ definiert worden, kann jeder Merkmalsmenge $\tilde{X} \subseteq \tilde{\mathcal{X}}$ auf folgende Weise eine Wahrscheinlichkeit zugeordnet werden:

$$\Pr[\dot{X}](\tilde{X}) := \Pr[\mathcal{G}](\dot{X}^{-1}(\tilde{X}))$$

$\Pr[\dot{X}]$ wird als *Wahrscheinlichkeitsverteilung* oder *Wahrscheinlichkeitsmaß* der Zufallsvariablen $\dot{X}$ bezeichnet, $\Pr[\dot{X}](\tilde{X})$ wird die Wahrscheinlichkeit genannt, mit der die Zufallsvariable $\dot{X}$ Werte in der Merkmalsmenge $\tilde{X}$ annimmt. Alternativ werden folgende Schreibweisen verwendet:

$$\Pr(\dot{X} \in \tilde{X}) := \Pr[\dot{X}](\tilde{X}) \quad \text{und} \quad \Pr(\dot{X} = x) := \Pr[\dot{X}](\{x\})$$

Wie bereits erwähnt wurde, gibt es in formaler Hinsicht keine Unterschiede zwischen Wahrscheinlichkeitsverteilungen von Zufallsvariablen und Häufigkeitsverteilungen statistischer Variablen. Infolgedessen können alle für statistische Variablen definierten Begriffe in formal identischer Weise für Zufallsvariablen definiert werden.

Gleichwohl ist es sinnvoll, Zufallsvariablen und statistische Variablen zu unterscheiden; dem dient der zur Notation von Zufallsvariablen verwendete Punkt. Es gibt insbesondere folgende Unterschiede:

- Während sich bei einer statistischen Variablen $X : \Omega \longrightarrow \tilde{\mathcal{X}}$ der Definitionsbereich Ω auf eine Menge realer oder fiktiver Objekte oder Situationen bezieht, denen durch die Variable bestimmte, als Daten erfassbare Merkmalswerte zugeordnet werden, bezieht sich der Definitionsbereich einer Zufallsvariablen $\dot{X}$ auf den Merkmalsraum eines Zufallsgenerators.
- Dem entspricht, dass sich statistische Variablen auf bei bestimmten Objekten bzw. in bestimmten Situationen realisierte oder als realisiert angenommene Merkmalswerte beziehen, dagegen Zufallsvariablen auf Verfahren oder Prozesse, durch die Situationen mit bestimmten Eigenschaften entstehen können oder entstanden sein könnten, wobei also eine modale Betrachtungsweise vorausgesetzt wird.
- Somit unterscheiden sich trotz ihrer formal gleichen Eigenschaften die Häufigkeitsverteilung $P[X]$ und die Wahrscheinlichkeitsverteilung $\Pr[\dot{X}]$. Während sich die Häufigkeitsverteilung auf einen (statistisch konstruierten) Sachverhalt bezieht, also auf Häufigkeiten der bei einer Referenzmenge Ω *realisierten* Merkmalswerte, bezieht sich die Wahrscheinlichkeitsverteilung einer Zufallsvariablen auf Eigenschaften eines Verfahrens, durch das Objekte oder Situationen mit bestimmten Eigenschaften entstehen können.

4. Stochastische Modellvariablen. Für viele Zwecke genügt es, Zufallsvariablen durch ihren Merkmalsraum und ihre Wahrscheinlichkeitsverteilung zu definieren. (Dem entspricht, dass zwei Zufallsvariablen als identisch angesehen werden, wenn ihre Merkmalsräume und Wahrscheinlichkeitsverteilungen identisch sind.) In gewisser Weise genügt diese Vorstellung auch zur begrifflichen Repräsentation von Zufallsgeneratoren. Denn um sich auf einen Zufallsgenerator $\mathcal{G}$ mit einem Merkmalsraum $\tilde{\mathcal{X}}$ zu beziehen, kann man auch eine Zufallsvariable $\dot{X} : \tilde{\mathcal{X}} \longrightarrow \tilde{\mathcal{X}}$ verwenden, die für alle $x \in \tilde{\mathcal{X}}$ durch $\dot{X}(x) := x$, also als identische Abbildung, definiert ist. Dann hat $\dot{X}$ die gleiche Wahrscheinlichkeitsverteilung wie $\mathcal{G}$.

Wenn man dynamische Modelle konstruieren möchte, ist jedoch die Vorstellung, dass die beteiligten stochastischen und nicht-stochastischen *Modellvariablen* ihre Werte durch reflektierbare Prozesse annehmen, von zentraler Bedeutung; denn die Aufgabe der Modellbildung besteht im Wesentlichen darin, Annahmen über diese Prozesse vorzunehmen und ihre Implikationen auszuarbeiten.

Soweit es sich bei den Modellvariablen um Zufallsvariablen handelt, ist es also erforderlich, sich auf Zufallsgeneratoren zu beziehen, durch deren Aktivierung ihre Werte erzeugt werden können. Für die praktische Durchführung können beliebige Geräte bzw. Verfahren eingesetzt werden. Auf sehr flexible Weise können auch Computer verwendet werden, die einen Zufallsgenerator $\bar{\mathcal{G}}$ zur Verfügung stellen, mit dem im Intervall $[0, 1]$ gleichverteilte Zufallszahlen erzeugt werden können. Sei nämlich $\dot{X}$ eine Zufallsvariable mit dem Merkmalsraum $\tilde{\mathcal{X}} = \{\tilde{x}_1, \ldots, \tilde{x}_m\}$ und einer Wahrscheinlichkeitsverteilung, die den Elementen von $\tilde{\mathcal{X}}$ die Wahrscheinlichkeiten $q_j := \Pr(\dot{X} = \tilde{x}_j)$ zuordnet. Dann liefert folgendes Verfahren Realisierungen dieser Zufallsvariablen:

$$\text{Aktivierung} \longrightarrow \boxed{\bar{\mathcal{G}}} \longrightarrow \epsilon \longrightarrow \tilde{x}_k$$

Hierbei ist ϵ die durch die Aktivierung von $\bar{\mathcal{G}}$ erzeugte Zufallszahl und $k = \min\{j \,|\, \epsilon \leq \sum_{i=1}^{j} q_i\}$. Wird das Verfahren n-mal wiederholt, entstehen n Werte einer statistischen Variablen $X : \Omega \longrightarrow \tilde{\mathcal{X}}$, deren Verteilung näherungsweise mit derjenigen von $\dot{X}$ übereinstimmt: $P[X] \approx \Pr[\dot{X}]$.

Oft genügt es bereits, Zufallsgeneratoren zu verwenden, die nur zwei Werte erzeugen können: eine 1 mit der Wahrscheinlichkeit p und eine 0 mit der Wahrscheinlichkeit $1-p$. Solche Zufallsgeneratorennen nennen wir *Ereignisgeneratoren* und verwenden für sie die Notation $\mathcal{G}_e[p]$.

5. Einfache stochastische Ablaufschemas. Wir beginnen mit zwei einfachen Varianten stochastischer Ablaufschemas. Beide gehen von einer diskreten Zeitachse $\mathcal{T}^* = \{0, 1, 2, 3, \ldots\}$ aus.

a) In einer ersten Variante besteht das Ablaufschema nur aus einer Folge von Zufallsvariablen $\dot{X}_t$, die für die Zeitstellen $t \in \mathcal{T}^*$ definiert sind. Das Ablaufschema kann also durch eine Menge

$$\{\dot{X}_t \,|\, t \in \mathcal{T}^*\} \tag{4.1}$$

repräsentiert werden. Es wird angenommen, dass alle Zufallsvariablen den gleichen Merkmalsraum haben.

b) In einer zweiten Variante besteht das Ablaufschema aus einer Menge

$$\{(\dot{X}_t, \ddot{Z}_t) \,|\, t \in \mathcal{T}^*\} \tag{4.2}$$

In jeder Zeitstelle t gibt es wie im ersten Fall eine Zufallsvariable $\dot{X}_t$, außerdem eine nicht-stochastische Modellvariable $\ddot{Z}_t$. Wie bei Zufallsvariablen wird auch für die nicht-stochastischen Modellvariablen angenommen, dass es jeweils einen (zu modellierenden) Prozess gibt, durch den bestimmte Werte der Variablen entstehen können. Stochastische und nicht-stochastische Modellvariablen ähneln sich also in ihrem modalen Charakter und unterscheiden sich grundsätzlich von statistischen Variablen. Um nicht-stochastische Modellvariablen sowohl von statistischen als auch von Zufallsvariablen zu unterscheiden, kennzeichnen wir sie durch zwei Punkte.

Es sei angemerkt, dass es sich sowohl bei den stochastischen als auch bei den nicht-stochastischen Modellvariablen um mehrdimensionale Variablen handeln kann.

6. Stochastische Ablaufschemas und Prozesse. Stochastische Ablaufschemas definieren zunächst nur einen Rahmen für mögliche Prozessabläufe. Bestimmte Prozesse entstehen erst dadurch, dass für die Zufallsvariablen $\dot{X}_t$, und ggf. für die nicht-stochastischen Modellvariablen $\ddot{Z}_t$, bestimmte Werte erzeugt werden. Als Ergebnis entstehen dann jeweils bestimmte Zeitreihen

$$x_0, x_1, x_2, \ldots \quad \text{oder} \quad (x_0, z_0), (x_1, z_1), (x_2, z_2), \ldots$$

wobei x_t und z_t die Werte sind, die für die Variablen $\dot{X}_t$ bzw. $\ddot{Z}_t$ in der Zeitstelle t realisiert wurden.

Offenbar kann man sich bei einem stochastischen Ablaufschema beliebig viele Realisierungen vorstellen, ggf. auch tatsächlich erzeugen. Bezieht man sich zum Beispiel auf n Realisierungen eines Ablaufschemas $\{\dot{X}_t \,|\, t \in \mathcal{T}^*\}$, können sie als ein diachron aggregierter statistischer Prozess aufgefasst werden: $X_t : \Omega \longrightarrow \tilde{\mathcal{X}}$. Für jede Zeitstelle t ist X_t eine statistische Variable mit der unveränderlichen Referenzmenge $\Omega = \{\omega_1, \ldots, \omega_n\}$, deren Elemente zur Identifizierung der Zeitreihen dienen, die als Realisierungen des Ablaufschemas entstanden sind. Bezieht man sich auf eine bestimmte Zeitreihe ω, wird durch $X_t(\omega)$ ihr Wert in der Zeitstelle t, also der in dieser Zeitstelle für diese Zeitreihe realisierte Wert der Zufallsvariablen $\dot{X}_t$, erfasst.

7. Regeln für stochastische Ablaufschemas. Zunächst ist ein stochastisches Ablaufschema nur ein formaler Rahmen für mögliche Prozesse, nicht bereits ein dynamisches Modell. Um zu einem dynamischen Modell zu gelangen, sind zusätzlich Annahmen darüber erforderlich, wie jeweils bestimmte

Werte der stochastischen und nicht-stochastischen Modellvariablen entstehen können. Setzen wir voraus, dass nomologische Modelle konstruiert werden sollen, müssen dafür Regeln formuliert werden.

Solche Regeln sind natürlich nur für *endogene* Modellvariablen erforderlich, deren Werte innerhalb des Modells bestimmt werden sollen. Davon zu unterscheiden sind *exogene* Modellvariablen, deren Werte vor der Generierung von Modellprozessen durch den Modellkonstrukteur festgelegt werden, so dass sie als Parameter des Ablaufschemas angesehen werden können. Exogene Modellvariablen sind stets nicht-stochastisch.

Zur Erzeugung von Werten endogener Modellvariablen können folgende Arten von Regeln in Betracht gezogen werden:

a) Um Werte einer stochastischen Modellvariablen $\dot{X}_t$ zu erzeugen, kann ein Zufallsgenerator $\mathcal{G}$ verwendet werden, dessen Wahrscheinlichkeitsverteilung vorab festgelegt worden ist.

b) Eine andere Möglichkeit besteht darin, die Auswahl eines Zufallsgenerators zur Erzeugung von Werten von $\dot{X}_t$ von der Prozesszeit und/oder von vorab (in einer früheren oder in der gleichen Zeitstelle) entstandenen Werten anderer Modellvariablen abhängig zu machen.

c) Werte nicht-stochastischer Modellvariablen $\ddot{Z}_t$ werden normalerweise durch deterministische Funktionen in Abhängigkeit von der Prozesszeit und/oder vorab erzeugten Werten anderer Modellvariablen gebildet. Dies setzt natürlich $t > 0$ voraus, da $\ddot{Z}_0$ stets eine exogene Modellvariable ist.

Kapitel 5

Funktionale Modelle

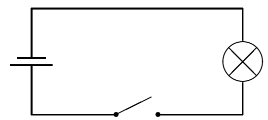
In diesem Kapitel beschäftigen wir uns in allgemeiner Weise mit *funktionalen Modellen*. Damit sind Modelle gemeint, die im Wesentlichen aus Modellvariablen bestehen, die durch Funktionen verknüpft sind. Diese

Funktionen können unterschiedlicher Art sein. Wir beginnen mit Modellen, bei denen einfache (deterministische) Funktionen verwendet werden, die die Merkmalsräume nicht-stochastischer Modellvariablen verknüpfen. Dann werden stochastische Varianten besprochen, bei denen Modellvariablen auch als stochastische Variablen konzipiert werden können. Der dritte Abschnitt beschäftigt sich mit einigen Fragen, die aus unterschiedlichen Konzeptions- und Deutungsmöglichkeiten stochastischer Modellvarianten resultieren. Schließlich werden im vierten Abschnitt indirekte Prognosefunktionen besprochen.

5.1 Deterministische Modelle

Ein *funktionales Modell* besteht aus einer Menge von *Modellvariablen*, die durch Funktionen miteinander verknüpft sind. Von einem *deterministischen* funktionalen Modell wird gesprochen, wenn das Modell keine stochastischen Variablen enthält und alle Funktionen deterministisch sind, d.h. in deterministischer Form die Wertebereiche der Variablen verknüpfen. In diesem Abschnitt beziehen wir uns nur auf deterministische Modelle.

1. Ein einfaches Beispiel. Zur Erläuterung betrachten wir ein einfaches Beispiel, das zunächst durch folgendes Bild illustriert werden kann:



In diesem Bild sind eine Stromquelle und eine Glühbirne durch einen Stromkreis verbunden, der durch einen Schalter geschlossen oder geöffnet werden kann. Je nachdem, ob der Schalter geschlossen oder geöffnet wird, leuchtet die Glühbirne oder nicht.

Für die in diesem Bild dargestellte Situation können unterschiedliche funktionale Modelle gebildet werden. Eine einfache Variante verwendet drei Modellvariablen:[1] Eine Variable $\ddot{Y}$, durch die erfasst wird, ob die Glühbirne leuchtet ($\ddot{Y} = 1$) oder nicht leuchtet ($\ddot{Y} = 0$); eine Variable $\ddot{X}$, durch die erfasst wird, ob der Schalter geschlossen ($\ddot{X} = 1$) oder nicht geschlossen ($\ddot{X} = 0$) ist; und eine Variable $\ddot{Z}$, durch die erfasst wird, ob die Batterie Strom abgeben kann ($\ddot{Z} = 1$) oder nicht ($\ddot{Z} = 0$). Außerdem wird eine Funktion gebildet, durch die die möglichen Werte von $\ddot{Y}$ von den Werten von $\ddot{X}$ und $\ddot{Z}$ abhängig gemacht werden. Folgende Tabelle definiert

[1] Wie bereits in Abschnitt 4.3 (§ 5) erläutert wurde, kennzeichnen wir nicht-stochastische Modellvariablen durch zwei Punkte, um sie von statistischen und von stochastischen bzw. Zufallsvariablen zu unterscheiden.

die Funktion:

$$\begin{array}{ccc} \ddot{X} & \ddot{Z} & \ddot{Y} \\ \hline 0 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{array} \tag{5.1}$$

Nur wenn die Batterie Strom abgeben kann und der Schalter geschlossen ist, leuchtet die Glühbirne.

2. Exogene und endogene Modellvariablen. Ein funktionales Modell besteht zunächst aus Variablen, wir nennen sie *Modellvariablen*, zu deren Definition folgende Angaben gemacht werden müssen:

– Es muss ein *Name der Variablen* angegeben und ihre Bedeutung erklärt werden. Zum Beispiel wird erklärt, dass mit einer Modellvariablen $\ddot{Y}$ erfasst werden soll, ob eine Glühbirne leuchtet oder nicht.

– Es muss der *Wertebereich* der Modellvariablen angegeben werden, d.h. die Menge ihrer möglichen Werte;[2] dabei setzen wir voraus, dass jede Modellvariable mindestens zwei unterschiedliche Werte annehmen kann. Zum Beispiel wird als Wertebereich der Variablen $\ddot{Y}$ die Menge $\{0, 1\}$ angegeben und erläutert, dass 1 bedeuten soll, dass die Glühbirne leuchtet, und dass 0 bedeuten soll, dass sie nicht leuchtet.

– Schließlich muss deutlich gemacht werden, wie bestimmte Werte einer Modellvariablen entstehen können.

Insbesondere die zuletzt genannte Angabe ist von wesentlicher Bedeutung, um Modellvariablen und somit auch funktionale Modelle zu verstehen.

In dieser Hinsicht kann man zunächst endogene und exogene Modellvariablen unterscheiden. *Endogene Modellvariablen* bekommen ihre Werte als Funktionen anderer Modellvariablen. In dem Beispiel aus § 1 ist $\ddot{Y}$ eine endogene Variable, deren Wert aus den Werten von $\ddot{X}$ und $\ddot{Z}$ resultiert. Modellvariablen, die nicht endogen sind, werden *exogene Modellvariablen* genannt. In unserem Beispiel sind $\ddot{X}$ und $\ddot{Z}$ exogene Modellvariablen.

3. Funktionen und nomologische Regeln. In einem funktionalen Modell sind die Modellvariablen durch Funktionen verknüpft. In einem deterministischen Modell handelt es sich um Funktionen der Form $f : \tilde{\mathcal{X}} \longrightarrow \tilde{\mathcal{Y}}$, wobei $\tilde{\mathcal{X}}$ der Wertebereich einer Variablen $\ddot{X}$ und $\tilde{\mathcal{Y}}$ der Wertebereich einer Variablen $\ddot{Y}$ ist. Jedem Element $x \in \tilde{\mathcal{X}}$ wird durch f *genau ein Element*

[2] In synonymer Bedeutung wird auch vom *Merkmalsraum* einer Modellvariablen gesprochen. Wie bei statistischen und Zufallsvariablen nehmen wir auch bei Modellvariablen an, dass es für ihre Wertebereiche stets eine numerische Repräsentation gibt. Wertebereiche von Modellvariablen sind also stets Mengen von Zahlen, für die eine Bedeutung vereinbart worden ist.

$f(x) \in \tilde{\mathcal{Y}}$ zugeordnet.[3] Dementsprechend kann man sagen, dass die Funktion f angibt, wie die Werte von $\ddot{Y}$ durch die Werte von $\ddot{X}$ bestimmt werden, oder in einer kurzen Formulierung: wie $\ddot{Y}$ von $\ddot{X}$ abhängt.

Solche Abhängigkeitsbeziehungen können graphisch durch Pfeile dargestellt werden. In diesem Beispiel: $\ddot{X} \longrightarrow \ddot{Y}$. Eine Variable kann natürlich von zwei oder mehr anderen Variablen simultan abhängig sein. Für das Beispiel aus § 1 kann folgende Darstellung verwendet werden:

$$\ddot{X} \longrightarrow \ddot{Y} \longleftarrow \ddot{Z}$$

Bei Darstellungen dieser Art muss jedoch beachtet werden, dass nicht jedem Pfeil eine separate Funktion entspricht; vielmehr beziehen sich alle Pfeile, die in einer Variablen münden, auf *eine* Funktion. In dem Beispiel handelt es sich um eine Funktion

$$g : \tilde{\mathcal{X}} \times \tilde{\mathcal{Z}} \longrightarrow \tilde{\mathcal{Y}},$$

die jedem Element $(x, z) \in \tilde{\mathcal{X}} \times \tilde{\mathcal{Z}}$ einen Wert $g(x, z) \in \tilde{\mathcal{Y}}$ zuordnet (so wie in Tabelle (5.1) in § 1 angegeben wurde). Wenn zwei Variablen durch einen Pfeil verknüpft sind, besagt dies also nur, dass es eine Funktion gibt, bei der eine Variable als abhängige, die andere als eine Argumentvariable auftritt.

Da der mathematische Funktionsbegriff für viele unterschiedliche Zwecke verwendet werden kann, sollte darauf geachtet werden, welche spezifische Bedeutung den bei der Konstruktion funktionaler Modelle verwendeten Funktionen zukommt: Es handelt sich um *nomologische Regeln*, die angeben, wie die endogenen Variablen des Modells in Abhängigkeit von anderen (endogenen oder exogenen) Variablen bestimmte Werte annehmen. Sie *bestimmen*, so könnte man auch sagen, die Prozesse, durch die die endogenen Variablen bestimmte Werte annehmen.

Natürlich gelten solche Regeln nur für das Modell, nicht auch unmittelbar für die Realität, auf die sich die Modellkonstruktion vielleicht beziehen soll. Zum Beispiel gilt für das in § 1 eingeführte Modell, dass die Glühbirne leuchtet, wenn die Batterie Strom abgeben kann und der Schalter geschlossen ist. Ob das auch bei einem realen System zutrifft, das man sich aus einer Glühbirne, einer Batterie, einem Schalter und einigen Drähten gebastelt hat, ist offenbar eine ganz andere Frage.

4. Funktionale Modelle als Ablaufschemas. Ein funktionales Modell kann auch als ein Ablaufschema betrachtet werden (vgl. Abschnitt 3.1, § 4), das festlegt, wie man von Werten der exogenen Variablen zu Werten der endogenen Variablen gelangt. In der graphischen Darstellung deuten die Pfeile die möglichen Prozesse an. Bei dynamischen funktionalen Modellen

[3] Dies ist gemeint, wenn von *deterministischen* Funktionen gesprochen wird, im Unterschied zu *stochastischen* Funktionen, durch die jedem Element von $\tilde{\mathcal{X}}$ eine Wahrscheinlichkeitsverteilung für die Elemente von $\tilde{\mathcal{Y}}$ zugeordnet wird.

entspricht der durch die Pfeile (Funktionen) definierten Richtung zugleich eine zeitliche Ordnung.

Unabhängig davon, ob sich die Funktionen temporal interpretieren lassen, kann man von wiederholbaren Prozessen sprechen. Zum Beispiel lässt sich bei dem Modell in § 1 der Schalter beliebig oft betätigen.

5. Formale Definition funktionaler Modelle. Ein funktionales Modell $\mathcal{M}$ kann durch $\mathcal{M} = (\mathcal{V}, \mathcal{F})$ definiert werden, wobei die Komponenten folgende Bedeutung haben:

$\mathcal{V}$ ist eine Menge von Modellvariablen.

$\mathcal{F}$ ist eine Menge von Funktionen, durch die Variablen in $\mathcal{V}$ verknüpft werden. Dabei gilt: Jede Variable in $\mathcal{V}$ tritt in höchstens einer Funktion als abhängige Variable, jedoch in mindestens einer Funktion entweder als abhängige oder als Argumentvariable auf. Außerdem wird angenommen, dass alle Variablen, die als Argumente verwendet werden, effektive Bedingungen für die jeweils abhängige Variable sind.[4]

Somit kann auch gesagt werden: Wenn eine Variable in einer Funktion als abhängige Variable auftritt, handelt es sich um eine endogene Variable, andernfalls um eine exogene Variable.

Jedem funktionalen Modell $\mathcal{M} = (\mathcal{V}, \mathcal{F})$ kann ein gerichteter Graph $\mathcal{G}(\mathcal{M}) = (\mathcal{V}, \mathcal{K})$ zugeordnet werden, wobei $\mathcal{V}$ die Knotenmenge und $\mathcal{K}$ die Menge der gerichteten Kanten (Pfeile) des Graphen bezeichnet. Die Variablen des Modells bilden die Knoten des Graphen, und zwei Variablen $V, V' \in \mathcal{V}$ sind genau dann durch eine gerichtete Kante (einen Pfeil) von V nach V' verbunden, wenn es in $\mathcal{F}$ eine Funktion gibt, bei der V als Argumentvariable und V' als abhängige Variable auftritt. Zu beachten ist also, dass meistens nicht jedem Pfeil eine Funktion entspricht, sondern dass alle Pfeile, die in der gleichen Variablen münden, zu *einer* Funktion in $\mathcal{F}$ gehören.

Der Graph eines funktionalen Modells zeigt die *Struktur des Modells*. Es handelt sich stets um einen einfachen Graphen ohne Schlingen. Wenn nicht ausdrücklich anders angegeben, soll auch angenommen werden, dass der Graph schwach zusammenhängend ist, also nicht in zwei oder mehr unabhängige Komponenten zerfällt, sowie nicht zyklisch ist. Zum Beispiel gibt es dann bei drei Modellvariablen (wenn man von Permutationen der Variablennamen absieht) folgende mögliche Strukturen:

$$\ddot{Z} \longleftarrow \ddot{X} \longrightarrow \ddot{Y} \qquad \ddot{X} \longrightarrow \ddot{Y} \longrightarrow \ddot{Z} \qquad \ddot{X} \longrightarrow \ddot{Y} \longleftarrow \ddot{Z}$$

Offenbar können unterschiedliche Modelle die gleiche Struktur aufweisen. Abhängig von den Wertebereichen der Modellvariablen kann die Anzahl unterschiedlicher Modelle auch sehr groß, ggf. auch unendlich groß werden. Wir verwenden folgende Sprechweisen:

[4] Was mit „effektiven Bedingungen“ gemeint ist, wird in § 9 definiert.

– Zwei Modelle $\mathcal{M} = (\mathcal{V}, \mathcal{F})$ und $\mathcal{M}' = (\mathcal{V}', \mathcal{F}')$ sind *strukturell identisch*, wenn ihre Graphen identisch sind.

6. Mögliche Werte und äquivalente Modelle. Weitere Überlegungen ergeben sich durch eine Bezugnahme auf die Werte, die die Variablen eines Modells annehmen können. Zunächst wird für jede Modellvariable durch ihren Wertebereich angegeben, welche Werte sie bei einer isolierten Betrachtung annehmen kann. Betrachtet man jedoch zwei oder mehr Modellvariablen gemeinsam, implizieren die Modellfunktionen Einschränkungen für die möglichen Wertekombinationen. Ist ein Modell $\mathcal{M} = (\mathcal{V}, \mathcal{F})$ gegeben und ist V eine (irgendwie geordnete) Auflistung von Variablen aus $\mathcal{V}$, soll $\mathrm{B}_{\mathcal{M}}[V]$ die Menge der im Rahmen des Modells möglichen *Belegungen* der Variablen in V mit bestimmten Werten bezeichnen.[5] Zum Beispiel findet man für das in § 1 angegebene Modell:

$$\mathrm{B}[\ddot{X}] = \tilde{\mathcal{X}}, \;\; \mathrm{B}[\ddot{Y}] = \tilde{\mathcal{Y}}, \;\; \mathrm{B}[\ddot{Z}] = \tilde{\mathcal{Z}}$$

Jedoch ist

$$\mathrm{B}[\ddot{X}, \ddot{Y}, \ddot{Z}] = \{(0,0,0), (0,0,1), (1,0,0), (1,1,1)\}$$

also nur eine Teilmenge von $\tilde{\mathcal{X}} \times \tilde{\mathcal{Y}} \times \tilde{\mathcal{Z}}$. Folgende Definition lässt sich anschließen:

– Zwei Modelle $\mathcal{M} = (\mathcal{V}, \mathcal{F})$ und $\mathcal{M}' = (\mathcal{V}', \mathcal{F}')$ sind *äquivalent*, wenn $\mathrm{B}_{\mathcal{M}}[\mathcal{V}] = \mathrm{B}_{\mathcal{M}'}[\mathcal{V}']$ ist.

Offenbar impliziert eine so definierte Äquivalenz, dass $\mathcal{V}$ und $\mathcal{V}'$ eineindeutig aufeinander abgebildet werden können; sie impliziert jedoch nicht, dass die Modelle strukturell identisch sind. Zur Illustration nehmen wir an, dass sich zwei Modelle auf die Variablen $\ddot{X}$, $\ddot{Y}$ und $\ddot{Z}$ beziehen und die Strukturen $\ddot{X} \longrightarrow \ddot{Y} \longrightarrow \ddot{Z}$ bzw. $\ddot{X} \longleftarrow \ddot{Y} \longrightarrow \ddot{Z}$ aufweisen. Gibt es beispielsweise für das erste Modell die Funktionen $y = 2\,x$ und $z = 4\,y$, kann man für das zweite Modell die Funktionen $x = 0.5\,y$ und $z = 4\,y$ wählen, so dass beide Modelle äquivalent sind.

7. Definierte und abgeleitete Funktionen. Wie bereits gesagt wurde, nehmen wir an, dass bei der Definition eines funktionalen Modells für jede endogene Variable genau eine Funktion angegeben wird, die ihre Abhängigkeit von anderen Modellvariablen festlegt. Diese Funktionen (also die Elemente von $\mathcal{F}$) nennen wir die *elementaren Funktionen* eines Modells. Darüber hinaus können oft weitere Funktionen abgeleitet werden. Hat ein Modell zum Beispiel die Struktur

$$\ddot{Z} \xrightarrow{f_{xz}} \ddot{X} \xrightarrow{f_{yx}} \ddot{Y}$$

[5] Innerhalb der eckigen Klammern können auch Mengen von Modellvariablen verwendet werden, für deren Elemente eine Reihenfolge vereinbart worden ist. Auch kann in der Notation der Verweis auf ein bestimmtes Modell entfallen, wenn der Bezug aus dem Kontext hervorgeht.

kann eine Funktion $f_{yz} : \tilde{\mathcal{Z}} \longrightarrow \tilde{\mathcal{Y}}$ abgeleitet werden, deren Werte durch $f_{yz}(z) = f_{yx}(f_{xz}(z))$ definiert sind. Entsprechend können auch bei komplizierteren Strukturen neue Funktionen gebildet werden. Werden zum Beispiel bei einem Modell mit der Struktur

$$\begin{tikzcd} \ddot{X} \arrow[rr] \arrow[dr] & & \ddot{Y} \\ & \ddot{Z} \arrow[ur] & \end{tikzcd}$$

die elementaren Funktionen durch $f_{yxz}(x,z)$ und $f_{zx}(x)$ bezeichnet, kann man durch $f_{yx}(x) = f_{yxz}(x, f_{zx}(x))$ eine neue Funktion definieren, in der Werte der Variablen $\ddot{Z}$ nicht mehr explizit vorkommen, die also einem einfachen Pfeil von $\ddot{X}$ nach $\ddot{Y}$ entspricht. Man erhält auf diese Weise eine Funktion, bei der als Argument nur noch eine exogene Variable vorkommt. Dies gilt offenbar generell: Stets kann für jede endogene Variable eine Funktion gebildet werden, bei der als Argumente nur noch diejenigen exogenen Variablen vorkommen, von denen aus ein gerichteter Pfad zu der betreffenden endogenen Variablen führt.[6]

8. Eindeutige und mengenwertige Funktionen. Für einige Überlegungen ist es zweckmäßig, auch *mengenwertige Funktionen* zu verwenden. Im Unterschied zu einer gewöhnlichen (deterministischen) Funktion wird durch eine mengenwertige Funktion zwischen Modellvariablen $\ddot{X}$ und $\ddot{Y}$ jedem Element des Wertebereichs von $\ddot{X}$ nicht ein Element, sondern eine Teilmenge des Wertebereichs von $\ddot{Y}$ zugeordnet.

Zur Illustration verwenden wir das Beispiel aus § 1. Obwohl es keine gewöhnliche Funktion $\ddot{Y} \longrightarrow (\ddot{X}, \ddot{Z})$ gibt, kann der Zusammenhang durch eine mengenwertige Funktion charakterisiert werden:

$$\begin{aligned} y = 0 &\longrightarrow \{(0,0), (0,1), (1,0)\} \\ y = 1 &\longrightarrow \{(1,1)\} \end{aligned}$$

Zur flexiblen Bezugnahme auf mengenwertige Funktionen, die innerhalb eines funktionalen Modells gebildet werden können, eignet sich eine Erweiterung der bereits eingeführten Notation für Belegungen:

– $\mathrm{B}_{\mathcal{M}}[\ddot{Y}|\ddot{X} = x]$ soll die Menge aller Werte y der Modellvariablen $\ddot{Y}$ bezeichnen, so dass es eine Belegung aller Modellvariablen gibt, bei der $\ddot{X}$ den Wert x und $\ddot{Y}$ den Wert y hat. (Sinngemäß können vor und nach dem Bedingungsstrich auch mehrere Modellvariablen stehen.)

Verweist $\mathcal{M}$ beispielsweise auf das Modell in § 1, ist $\mathrm{B}_{\mathcal{M}}[\ddot{X}, \ddot{Z}|\ddot{Y} = 0] = \{(0,0), (0,1), (1,0)\}$ und $\mathrm{B}_{\mathcal{M}}[\ddot{X}, \ddot{Z}|\ddot{Y} = 1] = \{(1,1)\}$. Also kann die Funktion $\ddot{Y} \longrightarrow (\ddot{X}, \ddot{Z})$ durch $y \longrightarrow \mathrm{B}_{\mathcal{M}}[\ddot{X}, \ddot{Z}|\ddot{Y} = y]$ dargestellt werden.

[6] Bezieht man sich nur auf diese Funktionen, wird auch von der *reduzierten Form* eines Modells oder von einem *reduzierten Modell* gesprochen.

9. Abhängigkeitsbeziehungen zwischen Variablen. Funktionale Modelle sollen helfen, Abhängigkeitsbeziehungen zwischen Variablen reflektierbar zu machen. Dafür ist es zweckmäßig, zwei unterschiedliche Abhängigkeitsbegriffe zu verwenden. Für die folgenden Definitionen beziehen wir uns auf ein Modell $\mathcal{M} = (\mathcal{V}, \mathcal{F})$ und zwei Variablen $\ddot{X}$ und $\ddot{Y}$ aus $\mathcal{V}$.

– $\ddot{Y}$ ist im Modell $\mathcal{M}$ von $\ddot{X}$ *abhängig*, wenn es Werte $x, x' \in \mathrm{B}_{\mathcal{M}}[\ddot{X}]$ gibt, so dass $\mathrm{B}_{\mathcal{M}}[\ddot{Y}|\ddot{X} = x] \neq \mathrm{B}_{\mathcal{M}}[\ddot{Y}|\ddot{X} = x']$ ist.

– $\ddot{Y}$ ist im Modell $\mathcal{M}$ von $\ddot{X}$ *funktional abhängig*, wenn $\ddot{Y}$ von $\ddot{X}$ abhängig ist und es einen gerichteten Pfad gibt, der von $\ddot{X}$ zu $\ddot{Y}$ führt. In diesem Fall verwenden wir auch die Redeweise, dass $\ddot{X}$ eine *effektive Bedingung* für $\ddot{Y}$ ist.

Man beachte, dass die Existenz eines gerichteten Pfads von $\ddot{X}$ nach $\ddot{Y}$ nicht ausreicht, um Abhängigkeit zu begründen. Wichtig ist auch, dass Abhängigkeit (im Unterschied zu funktionaler Abhängigkeit) in den meisten Anwendungsfällen symmetrisch ist, d.h. wenn $\ddot{Y}$ von $\ddot{X}$ abhängig ist, dann ist auch $\ddot{X}$ von $\ddot{Y}$ abhängig.[7] Somit kann auch ein symmetrischer Unabhängigkeitsbegriff verwendet werden:

– Zwei Variablen $\ddot{X}$ und $\ddot{Y}$ sind innerhalb eines Modells *unabhängig* (symbolisch durch $\ddot{X} \perp \ddot{Y}$ dargestellt), wenn innerhalb des Modells weder $\ddot{X}$ von $\ddot{Y}$ noch $\ddot{Y}$ von $\ddot{X}$ abhängig ist.

Verwendet man diese Definition, sind bei dem Modell aus § 1 $\ddot{X}$ und $\ddot{Z}$ unabhängig, dagegen $\ddot{X}$ und $\ddot{Y}$ sowie $\ddot{Z}$ und $\ddot{Y}$ abhängig.

Wenn zwei Variablen unabhängig sind, folgt offenbar, dass sie nicht abhängig und somit auch nicht funktional abhängig sind. Umgekehrt folgt jedoch daraus, dass zwei Variablen nicht funktional abhängig sind, nicht ihre Unabhängigkeit; denn sie können beide von einer dritten Variablen abhängig sein.

Es sei angemerkt, dass sich die Definitionen sinngemäß auch bei mehrdimensionalen Variablen verwenden lassen. Aus paarweiser Unabhängigkeit folgt jedoch nicht eine gemeinsame Unabhängigkeit; zum Beispiel kann man aus $\ddot{X} \perp \ddot{Z}$ und $\ddot{Y} \perp \ddot{Z}$ ohne weiteres nicht ableiten, dass auch $(\ddot{X}, \ddot{Y}) \perp \ddot{Z}$ gilt.[8]

[7] Am einfachsten sieht man das, wenn man sich $\mathrm{B}[\ddot{X}] \times \mathrm{B}[\ddot{Y}]$ als eine zweidimensionale Tabelle vorstellt. Allerdings muss vorausgesetzt werden, dass sowohl $\mathrm{B}_{\mathcal{M}}[\ddot{X}]$ als auch $\mathrm{B}_{\mathcal{M}}[\ddot{Y}]$ mindestens zwei Elemente enthalten.

[8] Man betrachte als Beispiel folgende Wertebelegungen:

$\ddot{X}$	$\ddot{Y}$	$\ddot{Z}$
0	0	0
0	1	0
1	1	0
0	1	1
1	0	1

Schließlich ist es in einigen Zusammenhängen nützlich, auch von *konditionaler Unabhängigkeit* sprechen zu können. Wir verwenden folgende Definitionen, wobei $\ddot{X}$, $\ddot{Y}$ und $\ddot{Z}$ drei Variablen eines Modells $\mathcal{M}$ sind:

- Im Modell $\mathcal{M}$ sind $\ddot{X}$ und $\ddot{Y}$ unter der Bedingung $\ddot{Z} = z$ voneinander unabhängig ($\ddot{X} \perp \ddot{Y} | \ddot{Z} = z$), wenn es keine Werte $x, x' \in \mathrm{B}_{\mathcal{M}}[\ddot{X}]$ gibt, so dass $\mathrm{B}_{\mathcal{M}}[\ddot{Y} | \ddot{X} = x, \ddot{Z} = z] \neq \mathrm{B}_{\mathcal{M}}[\ddot{Y} | \ddot{X} = x', \ddot{Z} = z]$ ist.
- Im Modell $\mathcal{M}$ sind $\ddot{X}$ und $\ddot{Y}$ unter der Bedingung $\ddot{Z}$ voneinander unabhängig ($\ddot{X} \perp \ddot{Y} | \ddot{Z}$), wenn $\ddot{X} \perp \ddot{Y} | \ddot{Z} = z$ für alle $z \in \mathrm{B}_{\mathcal{M}}[\ddot{Z}]$ gilt.

Verwendet man diese Definitionen, gilt zum Beispiel bei dem Modell aus § 1 zwar $\ddot{X} \perp \ddot{Y} | \ddot{Z} = 0$, nicht jedoch $\ddot{X} \perp \ddot{Y} | \ddot{Z} = 1$.

Bemerkenswert ist, dass aus konditionaler Unabhängigkeit nicht einfache Unabhängigkeit folgt. Zur Illustration betrachten wir ein Modell $\ddot{X} \longleftarrow \ddot{Z} \longrightarrow \ddot{Y}$, bei dem $\ddot{X}$ und $\ddot{Y}$ funktional von einer Variablen $\ddot{Z}$ abhängig sind. Es gilt dann $\ddot{X} \perp \ddot{Y} | \ddot{Z}$, nicht unbedingt auch $\ddot{X} \perp \ddot{Y}$.[9] Wie das in Anmerkung 8 angegebene Beispiel zeigt, folgt umgekehrt aus $\ddot{X} \perp \ddot{Y}$ auch nicht immer eine bedingte Unabhängigkeit $\ddot{X} \perp \ddot{Y} \,|\, \ddot{Z}$.

10. Prognosemöglichkeiten und effektive Bedingungen. Wenn in einem Modell $\mathcal{M}$ zwei Variablen $\ddot{X}$ und $\ddot{Y}$ abhängig sind, kann man sie wechselseitig zur Prognose von Werten der jeweils anderen Variablen verwenden. Zwei Möglichkeiten können unterschieden werden:

- $\ddot{Y}$ ist in einem Modell $\mathcal{M}$ durch $\ddot{X} = x$ *eindeutig prognostizierbar*, wenn $\mathrm{B}_{\mathcal{M}}[\ddot{Y} | \ddot{X} = x]$ genau ein Element enthält.
- $\ddot{Y}$ ist in einem Modell $\mathcal{M}$ durch $\ddot{X} = x$ *unbestimmt prognostizierbar*, wenn $\mathrm{B}_{\mathcal{M}}[\ddot{Y} | \ddot{X} = x]$ zwei oder mehr Elemente enthält, jedoch eine echte Teilmenge von $\mathrm{B}_{\mathcal{M}}[\ddot{Y}]$ ist.

Orientiert man sich an diesen Definitionen, setzt Prognostizierbarkeit der Werte einer Variablen $\ddot{Y}$ durch Werte einer Variablen $\ddot{X}$ nicht voraus, dass $\ddot{Y}$ von $\ddot{X}$ funktional abhängig ist, so dass man davon reden könnte, dass $\ddot{X}$ eine effektive Bedingung für $\ddot{Y}$ ist. Zum Beispiel sind in dem Modell in § 1 die Werte von $\ddot{X}$ und $\ddot{Z}$ durch $\ddot{Y} = 1$ (nicht jedoch durch $\ddot{Y} = 0$) eindeutig prognostizierbar, obwohl $\ddot{Y}$ weder für $\ddot{X}$ noch für $\ddot{Z}$ eine effektive Bedingung ist.

11. Mehrdimensionale Variablen und Constraints. Die Variablen eines

[9]Man betrachte beispielsweise die Wertebelegungen

$\ddot{Z}$	$\ddot{X}$	$\ddot{Y}$
0	0	0
1	1	0
2	0	1
3	1	1

$\ddot{Z}$	$\ddot{X}$	$\ddot{Y}$
0	0	0
1	1	1
2	0	1
3	1	1

In der linken Variante sind $\ddot{X}$ und $\ddot{Y}$ unabhängig, in der rechten Variante sind sie abhängig.

funktionalen Modells können ein- oder mehrdimensional sein. Ist $\ddot{X}$ eine Modellvariable, kann man beispielsweise annehmen, dass sie aus m Komponenten besteht, also die Form $\ddot{X} = (\ddot{X}_1, \ldots, \ddot{X}_m)$ hat. Ebenso kann man bereits definierte Variablen zusammenfassen, z.B. aus $\ddot{X}$ und $\ddot{Y}$ eine zweidimensionale Variable $(\ddot{X}, \ddot{Y})$ mit dem Wertebereich $\tilde{\mathcal{X}} \times \tilde{\mathcal{Y}}$ bilden.

Die Verwendung mehrdimensionaler Variablen ist insbesondere dann zweckmäßig, wenn zur Formulierung eines funktionalen Modells *Constraints* für die möglichen Wertekombinationen bei zwei oder mehr Modellvariablen definiert werden. Sind etwa $\ddot{X}$ und $\ddot{Y}$ zwei Variablen mit den Wertebereichen $\tilde{\mathcal{X}} = \tilde{\mathcal{Y}} = \{1, \ldots, 5\}$, könnten die möglichen Wertekombinationen durch die Bedingung $x + y \leq 6$ eingeschränkt werden. Offenbar sind dann $\ddot{X}$ und $\ddot{Y}$ abhängig; aber diese Abhängigkeit kann nicht durch funktionale Abhängigkeiten zwischen den beiden Variablen formuliert werden.[10] Die Bedingung, die diese Abhängigkeit erzeugt, kann jedoch auf einfache Weise dadurch erfasst werden, dass man die beiden Variablen zu einer zweidimensionalen Variablen $(\ddot{X}, \ddot{Y})$ zusammenfasst und für sie den Wertebereich $\{(x, y) \in \tilde{\mathcal{X}} \times \tilde{\mathcal{Y}} \,|\, x + y \leq 6\}$ angibt.

Offenbar können solche Constraints auch für exogene Variablen definiert werden. Exogene Variablen sind also nur dann voneinander unabhängig, wenn sie nicht durch Constraints abhängig gemacht werden. Sie sind jedoch niemals funktional voneinander abhängig, weil nur endogene Variablen von anderen Variablen funktional abhängig sein können.

12. Bedingungen für Unabhängigkeit. Die Existenz von Constraints ist insbesondere für die Frage bedeutsam, ob abhängige Variablen durch Konditionierung unabhängig gemacht werden können. In einem deterministischen Modell ohne Constraints ist das stets der Fall. Denn wenn $\ddot{X}$ und $\ddot{Y}$ zwei abhängige (per Voraussetzung endogene und nicht funktional voneinander abhängige) Modellvariablen sind und die Menge der Variablen, von denen aus ein Pfad zu $\ddot{X}$ oder $\ddot{Y}$ führt, aus $\ddot{Z}_1, \ldots, \ddot{Z}_m$ besteht, gilt stets: $\ddot{X} \perp \ddot{Y} \,|\, \ddot{Z}_1, \ldots, \ddot{Z}_m$.

Diese Feststellung, die in der Literatur oft als *deterministische Markow-Bedingung* bezeichnet wird, gilt jedoch nicht allgemein, wenn es Constraints gibt. Dann können nicht nur exogene Variablen abhängig sein, sondern es gibt auch nicht immer Modellvariablen, die als Bedingungen einer konditionalen Unabhängigkeit von $\ddot{X}$ und $\ddot{Y}$ verwendet werden könnten. Zwar kann stets eine Variable $\ddot{Z}$ konstruiert werden, so dass $\ddot{X} \perp \ddot{Y} \,|\, \ddot{Z}$ gilt.[11] Es handelt sich dann aber um eine fiktive Variable, die nicht zu den Variablen des vorausgesetzten Modells gehört.

Diese Überlegung hat auch Konsequenzen für folgendes *Prinzip der*

[10]Von Constraints soll in diesem Text generell nur dann gesprochen werden, wenn die Abhängigkeiten nicht durch Funktionen formuliert werden können.

[11]Es genügt, für jedes Paar von Werten $(x, y) \in \mathrm{B}[\ddot{X}, \ddot{Y}]$ einen unterschiedlichen Wert für $\ddot{Z}$ anzunehmen.

gemeinsamen Abhängigkeit: Wenn in einem funktionalen Modell zwei Modellvariablen $\ddot{X}$ und $\ddot{Y}$ abhängig sind, ist entweder $\ddot{X}$ funktional abhängig von $\ddot{Y}$ oder $\ddot{Y}$ funktional abhängig von $\ddot{X}$ oder es gibt eine weitere Variable $\ddot{Z}$, von der sowohl $\ddot{X}$ als auch $\ddot{Y}$ funktional abhängig sind, so dass $\ddot{X} \perp \ddot{Y} \mid \ddot{Z}$ gilt. – Zwar kann eine solche Variable $\ddot{Z}$ immer konstruiert werden, es ist aber nicht gewährleistet, dass eine solche Variable (oder Menge von Variablen) innerhalb des jeweils gegebenen Modells existiert, wenn bei der Definition des Modells Constraints verwendet werden.

5.2 Modelle mit stochastischen Variablen

In diesem Abschnitt beschäftigen wir uns mit funktionalen Modellen, die außer deterministischen Modellvariablen auch stochastische Variablen enthalten.

1. Deterministische und stochastische Funktionen. Werden zur Konstruktion eines funktionalen Modells sowohl deterministische als auch stochastische Variablen verwendet, müssen zwei Arten von Funktionen unterschieden werden:

a) Einerseits *deterministische Funktionen*, die die Wertebereiche von Modellvariablen verknüpfen. Folgende Fälle sind möglich:

$$\ddot{X} \longrightarrow \ddot{Y} \quad \text{und} \quad \dot{X} \longrightarrow \dot{Y}$$

Ist $\tilde{\mathcal{X}}$ der Wertebereich von $\ddot{X}$ bzw. $\dot{X}$ und ist $\tilde{\mathcal{Y}}$ der Wertebereich von $\ddot{Y}$ bzw. $\dot{Y}$, wird in beiden Fällen durch eine deterministische Funktion jedem Wert in $\tilde{\mathcal{X}}$ genau ein Wert in $\tilde{\mathcal{Y}}$ zugeordnet.[12] Im zweiten Fall erhält man durch

$$\Pr(\dot{Y} = y) \;=\; \textstyle\sum_{x \in f^{-1}(\{y\})} \Pr(\dot{X} = x)$$

die Verteilung von $\dot{Y}$ aus der Verteilung von $\dot{X}$ (wobei f der Name der Funktion $f : \tilde{\mathcal{X}} \longrightarrow \tilde{\mathcal{Y}}$ ist).

b) Andererseits können Modellvariablen durch *stochastische Funktionen* verknüpft werden. Folgende Fälle sind möglich:

$$\ddot{X} \twoheadrightarrow \dot{Y} \quad \text{und} \quad \dot{X} \twoheadrightarrow \dot{Y}$$

Ist $\tilde{\mathcal{X}}$ der Wertebereich von $\ddot{X}$ bzw. $\dot{X}$, wird jedem Wert $x \in \tilde{\mathcal{X}}$ im ersten Fall eine bedingte Verteilung $\Pr[\dot{Y}|\ddot{X} = x]$, im zweiten Fall eine bedingte Verteilung $\Pr[\dot{Y}|\dot{X} = x]$ zugeordnet. Orientiert man sich an den in Abschnitt 1.3 eingeführten Definitionen, handelt es sich um allgemeine Regressionsfunktionen.[13]

[12] Offenbar sind Verknüpfungen der Form $\ddot{X} \longrightarrow \dot{Y}$ und $\dot{X} \longrightarrow \ddot{Y}$ nicht möglich.

Da es sich um unterschiedliche Arten von Funktionen handelt, verwenden wir unterschiedliche Pfeile: $\longrightarrow$ für deterministische und $\twoheadrightarrow$ für stochastische Funktionen.

2. Ein einfaches technisches Beispiel. Um die Begriffsbildungen zu illustrieren, beziehen wir uns auf einen Toaster, bei dem die Glühdauer durch einen Hebel eingestellt werden kann. Die Stellung des Hebels wird durch eine exogene Modellvariable $\ddot{X}$, die Glühdauer durch eine endogene stochastische Variable $\dot{Y}$ erfasst. Die Variablen werden durch eine stochastische Funktion verknüpft: $\ddot{X} \twoheadrightarrow \dot{Y}$. Jedem möglichen Wert von $\ddot{X}$, also jeder möglichen Hebelstellung $x \in \tilde{\mathcal{X}}$, wird eine Verteilung $\Pr[\dot{Y}|\ddot{X} = x]$ der Glühdauervariablen zugeordnet. Wenn man sich nur dafür interessiert, wie die durchschnittliche Glühdauer von der Hebelstellung abhängt, kann der Zusammenhang auch durch eine spezielle Regressionsfunktion $x \longrightarrow \mathrm{E}(\dot{Y}|\ddot{X} = x)$ dargestellt werden.

3. Abhängige und unabhängige Variablen. Bei stochastischen Modellen kann die Vorstellung, dass zwei Variablen abhängig oder unabhängig sein können, auf unterschiedliche Weisen definiert werden. Zur Erläuterung beziehen wir uns auf ein Modell $\mathcal{M} = (\mathcal{V}, \mathcal{F})$ und zwei Variablen $\ddot{X}$ und $\dot{Y}$ (wenn beide Variablen stochastisch sind, können vollständig analoge Definitionen verwendet werden).

– Einerseits können die Begriffsbildungen übernommen werden, die in Abschnitt 5.1 (§ 9) definiert worden sind, insbesondere: $\ddot{X}$ und $\dot{Y}$ sind im Modell $\mathcal{M}$ *unabhängig* (symbolisch: $\ddot{X} \perp \dot{Y}$), wenn für alle Werte $x, x' \in \mathrm{B}_{\mathcal{M}}[\ddot{X}]$ gilt: $\mathrm{B}_{\mathcal{M}}[\dot{Y}|\ddot{X} = x] = \mathrm{B}_{\mathcal{M}}[\dot{Y}|\ddot{X} = x']$ ist; andernfalls sind $\ddot{X}$ und $\dot{Y}$ *abhängig*.

– Andererseits kann der Umstand ausgenutzt werden, dass mindestens eine stochastische Variable beteiligt ist, und folgende Definition verwendet werden: $\ddot{X}$ und $\dot{Y}$ sind im Modell $\mathcal{M}$ *stochastisch unabhängig* (symbolisch: $\ddot{X} \perp\!\!\!\perp \dot{Y}$), wenn es keine Werte $x, x' \in \mathrm{B}_{\mathcal{M}}[\ddot{X}]$ gibt, so dass $\Pr[\dot{Y}|\ddot{X} = x] \neq \Pr[\dot{Y}|\ddot{X} = x']$ ist; andernfalls sind $\ddot{X}$ und $\dot{Y}$ *stochastisch abhängig*.[14] Äquivalente Definitionen stochastischer Unabhängig-

[13] Offenbar ist es für diese Funktionen irrelevant, ob die Variablen, deren Werte als Bedingungen verwendet werden, deterministisch oder stochastisch sind. Somit können Bedingungen für Wahrscheinlichkeitsverteilungen auch mithilfe nicht-stochastischer Variablen formuliert werden können. Es gibt dann allerdings keine gemeinsame Verteilung. Aber auch dann, wenn als Bedingung für die Verteilung von $\dot{Y}$ eine stochastische Variable $\dot{X}$ verwendet wird, entsteht eine gemeinsame Verteilung erst dann, wenn auch eine Verteilung für $\dot{X}$ gegeben ist. Die stochastische Funktion $x \longrightarrow \Pr[\dot{Y}|\dot{X} = x]$ ist vollständig unabhängig von Annahmen über die Verteilung von $\dot{X}$ und sie setzt insbesondere nicht voraus, dass es für alle Werte $x \in \tilde{\mathcal{X}}$ eine positive Wahrscheinlichkeit gibt.

[14] Da $\ddot{X}$ eine deterministische Variable ist, können auch asymmetrische Sprechweisen verwendet werden: $\dot{Y}$ ist (einseitig) von $\ddot{X}$ stochastisch abhängig oder unabhängig.

keit sind: Für alle $(x,y) \in \mathrm{B}_{\mathcal{M}}[\ddot{X},\dot{Y}]$: $\Pr(\dot{Y}=y|\ddot{X}=x) = \Pr(\dot{Y}=y)$; und bei zwei stochastischen Variablen: Für alle $(x,y) \in \mathrm{B}_{\mathcal{M}}[\dot{X},\dot{Y}]$: $\Pr(\dot{Y}=y,\dot{X}=x) = \Pr(\dot{Y}=y)\Pr(\dot{X}=x)$.

Beide Abhängigkeits- bzw. Unabhängigkeitsbegriffe sind symmetrisch (wie in Abschnitt 5.1 (§ 9) kann auch ein nicht-symmetrischer Begriff der funktionalen Abhängigkeit definiert werden); und es gilt: Wenn zwei Variablen abhängig sind, sind sie auch stochastisch abhängig, und umgekehrt: wenn sie stochastisch unabhängig sind, sind sie auch unabhängig; aber daraus, dass zwei Variablen stochastisch abhängig sind, folgt nicht, dass sie abhängig sind.

Wie im deterministischen Fall folgt aus $\dot{X} \perp\!\!\!\perp \dot{Y}|\dot{Z}$ nicht die einfache Unabhängigkeit $\dot{X} \perp\!\!\!\perp \dot{Y}$; umgekehrt folgt aus einfacher nicht eine konditionale Unabhängigkeit; und ebenfalls wie im deterministischen Fall folgt aus paarweiser stochastischer Unabhängigkeit nicht eine gemeinsame stochastische Unabhängigkeit.

4. Notationen für stochastische Modelle. Zur formalen Repräsentation stochastischer Modelle kann wie bei deterministischen Modellen die Notation $\mathcal{M} = (\mathcal{V},\mathcal{F})$ verwendet werden. Natürlich muss bei den Elementen von $\mathcal{V}$ angegeben werden, ob es sich um deterministische oder stochastische Variablen handelt, und bei den Elementen von $\mathcal{F}$ muss zwischen deterministischen und stochastischen Funktionen unterschieden werden.

Es wird angenommen, dass ein stochastisches Modell mindestens eine endogene stochastische Variable enthält. Wir sprechen von stochastischen Modellen i.e.S., wenn alle endogenen Variablen stochastisch sind.

Exogene Modellvariablen können deterministisch oder stochastisch sein. Während Annahmen über bestimmte Werte exogener deterministischer Variablen nicht zur Modelldefinition gehören, betrachten wir Annahmen über die Verteilungen exogener stochastischer Variablen als Bestandteil der Modelldefinition. Gibt es mehrere exogene stochastische Variablen, muss ihre gemeinsame Verteilung angegeben werden. Es wird auch angenommen, dass stochastische exogene Variablen nicht durch Contraints mit deterministischen exogenen Variablen verknüpft sind, so dass sie von diesen stochastisch unabhängig sind.[15]

Wie deterministische Modelle können auch funktionale Modelle, die stochastische Variablen enthalten, durch gerichtete Graphen dargestellt werden. Jeder Knoten repräsentiert eine deterministische oder stochastische Modellvariable, und jede Kante gehört zu einer deterministischen ($\longrightarrow$) oder stochastischen ($\longrightarrow\!\!\!\!\!\rightarrow$) Funktion. Wie bei deterministischen Modellen verwenden wir folgende Sprechweise:

– Zwei Modelle $\mathcal{M} = (\mathcal{V},\mathcal{F})$ und $\mathcal{M}' = (\mathcal{V}',\mathcal{F}')$ sind *strukturell identisch*, wenn ihre Graphen identisch sind.

[15] Wenn Modellvariablen von deterministischen exogenen Variablen stochastisch abhängig sind, sind sie infolgedessen endogen; man vgl. auch Abschnitt 5.3 (§ 6).

5. Abgeschlossene stochastische Modelle. Die allgemeine Definition lässt auch Modelle zu, in denen alle exogenen Variablen (und infolgedessen auch alle endogenen Variablen) stochastisch sind. Wir nennen sie *abgeschlossene stochastische Modelle*. Ein einfaches Beispiel ist

$$\dot{X} \longrightarrow\!\!\!\!\!\rightarrow \dot{Y}$$

Bei abgeschlossenen stochastischen Modellen ist es möglich, eine gemeinsame Verteilung für alle Modellvariablen zu bilden. In dem Beispiel:

$$\Pr(\dot{X}=x,\dot{Y}=y) = \Pr(\dot{Y}=y|\dot{X}=x)\Pr(\dot{X}=x)$$

Außerdem kann, gewissermaßen als reduzierte Form eines abgeschlossenen stochastischen Modells, eine gemeinsame Verteilung aller endogenen Variablen gebildet werden; in dem Beispiel:

$$\Pr(\dot{Y}=y) = \textstyle\sum_x \Pr(\dot{X}=x,\dot{Y}=y)$$

In der Literatur werden stochastische Modelle oftmals als abgeschlossene Modelle konzipiert. Viele Autoren betrachten ausschließlich stochastische Modellvariablen und gehen von einer gemeinsamen Verteilung für alle Variablen eines Modells aus.[16] Dies hängt vermutlich damit zusammen, dass stochastische Modelle oft von vornherein auch als (Sampling-) Modelle für bestimmte Daten aufgefasst werden.[17] Wir werden dieser Betrachtungsweise nicht folgen, da ein funktionales Modell zunächst vollständig unabhängig von irgendwelchen Daten definiert werden muss. Ob bzw. unter welchen Umständen es sinnvoll sein kann, exogene Variablen als stochastische Variablen zu konzipieren, wird in Abschnitt 5.3 besprochen.

6. Beziehungen durch stochastische Funktionen. Bei deterministischen Modellen beziehen sich Fragestellungen darauf, wie die Werte der endogenen Variablen von den Werten der exogenen Variablen abhängen. Bei einem stochastischen Modell muss die Fragestellung verändert werden, da durch das Modell nur die Verteilung der stochastischen endogenen Variablen bestimmt wird. Dafür gibt es unterschiedliche Varianten. Zur Verdeutlichung beziehen wir uns auf ein stochastisches Modell i.e.S., das folgende (ggf. mehrdimensionalen) Modellvariablen enthält: eine deterministische exogene Variablen $\ddot{X}$, eine stochastische exogene Variablen $\dot{Z}$ und eine stochastische endogene Variablen $\dot{Y}$. Mehrere unterschiedliche Modellstrukturen sind möglich.

[16] Man vgl. beispielsweise Spirtes, Glymour und Scheines (1993), Cooper (1999); Greenland, Pearl und Robins (1999); Robins (1999); Pearl (2000); Woodward (2001: 41).

[17] Aus dieser Auffassung folgt dann die oft diskutierte Frage, wie und unter welchen Voraussetzungen man von gemeinsamen Verteilungen zu kausal interpretierbaren Beziehungen zwischen Variablen gelangen kann. Einen anderen Gedankengang, der nicht von abgeschlossenen Modellen ausgeht, verfolgen wir in Kapitel 6.

Zunächst kann man an ein Modell der Form $\ddot{X} \longrightarrow \dot{Y} \longleftarrow \dot{Z}$ denken, bei dem die Variablen durch eine deterministische Funktion $y = f(x, z)$ verbunden sind. Die endogene Variable $\dot{Y}$ ist in diesem Fall eine stochastische Variable, weil und insofern sie von der exogenen stochastischen Variablen $\dot{Z}$ abhängig ist. Berücksichtigt man nun, dass das Modell eine Annahme über die Verteilung von $\dot{Z}$ impliziert, gilt das Interesse nicht der deterministischen Funktion f, sondern einer aus dem Modell ableitbaren stochastischen Funktion:[18]

$$x \longrightarrow \Pr(\dot{Y} = y | \ddot{X} = x) = \sum_{z: f(x,z)=y} \Pr(\dot{Z} = z) \tag{5.2}$$

Diese stochastische Funktion zeigt, wie die Verteilung von $\dot{Y}$ von Werten der exogenen Variablen $\ddot{X}$ (über die im Rahmen des Modells unterschiedliche Annahmen getroffen werden können) abhängt.

Eine andere mögliche Modellstruktur ist $\ddot{X} \twoheadrightarrow \dot{Y} \twoheadleftarrow \dot{Z}$. In diesem Fall ist $\dot{Y}$ eine stochastische Variable, deren Verteilung von den Werten von $\ddot{X}$ und $\dot{Z}$ abhängt. Dem entspricht, dass durch das Modell eine stochastische Funktion $(x, z) \longrightarrow \Pr[\dot{Y} | \ddot{X} = x, \dot{Z} = z]$ vorgegeben wird, aus der sich dann analog zu (5.2), wiederum unter Ausnutzung der durch das Modell vorgegebenen Unabhängigkeit von $\dot{Z}$ und $\ddot{X}$, eine stochastische Funktion

$$x \longrightarrow \Pr[\dot{Y} | \ddot{X} = x] = \sum_{z \in \tilde{\mathcal{Z}}} \Pr[\dot{Y} | \ddot{X} = x, \dot{Z} = z] \Pr(\dot{Z} = z) \tag{5.3}$$

ableiten lässt. Im Unterschied zum ersten Fall überlagern bzw. mischen sich jetzt die Verteilungen von $\dot{Z}$ und $\dot{Y}$.

In beiden Fällen wird durch das vorausgesetzte Modell angenommen, dass $\dot{Z}$ und $\ddot{X}$ stochastisch unabhängig sind. Wenn das nicht der Fall ist, gibt es zusätzlich eine Abhängigkeit $\ddot{X} \twoheadrightarrow \dot{Z}$, also eine stochastische Funktion $x \longrightarrow \Pr[\dot{Z} | \ddot{X} = x]$. Dann ist zunächst bemerkenswert, dass $\dot{Z}$ dadurch zu einer endogenen Variablen wird. Analog zu (5.3) kann jedoch eine einfache stochastische Funktion zwischen $\ddot{X}$ und $\dot{Y}$ gebildet werden (es genügt, auf der rechten Seite die bedingte Verteilung $\Pr(\dot{Z} = z | \ddot{X} = x)$ zu verwenden).

7. Verkettung von Funktionen. In Modellen, in denen ein Pfad über mehrere Variablen verläuft, können die Funktionen verkettet werden. Man kann folgende Fälle unterscheiden.

- $\dot{Z} \longrightarrow \dot{X} \longrightarrow \dot{Y}$. Entsprechen den Pfeilen die Funktionen $y = f(x)$ und $x = g(z)$, kann man sie unmittelbar verketten: $y = f(g(z))$.
- $\dot{Z} \longrightarrow \dot{X} \twoheadrightarrow \dot{Y}$. Man kann dann eine stochastische Funktion bilden, die von $\dot{Z}$ zu $\dot{Y}$ führt: $z \longrightarrow \Pr[\dot{Y} | \dot{Z} = z] = \Pr[\dot{Y} | \dot{X} = g(z)]$.
- $\dot{Z} \twoheadrightarrow \dot{X} \longrightarrow \dot{Y}$. Eine stochastische Funktion, die von $\dot{Z}$ zu $\dot{Y}$ führt, hat die Form $z \longrightarrow \Pr(\dot{Y} = y | \dot{Z} = z) = \sum_{x \in f^{-1}(\{y\})} \Pr(\dot{X} = x | \dot{Z} = z)$.

[18] Die Modellstruktur impliziert, dass $\dot{Z}$ von $\ddot{X}$ stochastisch unabhängig ist.

- $\dot{Z} \twoheadrightarrow \dot{X} \twoheadrightarrow \dot{Y}$. Eine stochastische Funktion, die von $\dot{Z}$ zu $\dot{Y}$ führt, kann dann zunächst in folgender Form geschrieben werden:

$$z \longrightarrow \Pr(\dot{Y} = y | \dot{Z} = z) = \sum_{x \in \tilde{\mathcal{X}}} \Pr(\dot{Y} = y, \dot{X} = x | \dot{Z} = z) =$$
$$\sum_{x \in \tilde{\mathcal{X}}} \Pr(\dot{Y} = y | \dot{X} = x, \dot{Z} = z) \Pr(\dot{X} = x | \dot{Z} = z)$$

Es stellt sich die Frage, ob $\dot{Y}$ und $\dot{Z}$ unter der Bedingung, dass ein Wert von $\dot{X}$ gegeben ist, stochastisch unabhängig sind ($\dot{Y} \perp\!\!\!\perp \dot{Z} \,|\, \dot{X}$). Geht man von einer beliebigen Zufallsvariablen $(\dot{Y}, \dot{X}, \dot{Z})$ aus, ist das nicht immer der Fall. Die angegebenen vier Modellvarianten implizieren jedoch diese konditionale Unabhängigkeit. In den ersten drei Fällen kann sie unmittelbar bewiesen werden; im letzten Fall handelt es sich um eine durch die Modellstruktur gesetzte Annahme, so dass sich auch die dort angegebene stochastische Funktion vereinfacht:

$$\Pr[\dot{Y} | \dot{Z} = z] = \sum_{x \in \tilde{\mathcal{X}}} \Pr[\dot{Y} = y | \dot{X} = x] \Pr(\dot{X} = x | \dot{Z} = z)$$

8. Mehrere endogene stochastische Variablen. Jetzt betrachten wir unterschiedliche Modellstrukturen, wenn in einem Modell zwei endogene stochastische Variablen $\dot{Y}_1$ und $\dot{Y}_2$ explizit unterschieden werden, entweder separat oder als Komponenten einer zweidimensionalen Variablen $(\dot{Y}_1, \dot{Y}_2)$.

Eine mögliche Modellstruktur ist $\ddot{X} \twoheadrightarrow (\dot{Y}_1, \dot{Y}_2)$.[19] Durch das Modell wird eine stochastische Funktion $x \longrightarrow \Pr[\dot{Y}_1, \dot{Y}_2 | \ddot{X} = x]$ vorgegeben. Offenbar können $\dot{Y}_1$ und $\dot{Y}_2$ abhängig oder unabhängig sein. Dies gilt auch dann, wenn durch ein Modell $\dot{Y}_1 \twoheadleftarrow \ddot{X} \twoheadrightarrow \dot{Y}_2$ zwei separate stochastische Funktionen definiert werden.

Wenn es sich bei der exogenen Variablen um eine stochastische Variable handelt, wir nennen sie $\dot{Z}$, gibt es zwei Möglichkeiten. Bei einem Modell der Form $\dot{Z} \twoheadrightarrow (\dot{Y}_1, \dot{Y}_2)$ können $\dot{Y}_1$ und $\dot{Y}_2$ wie im ersten Fall abhängig oder unabhängig sein. Wenn jedoch das Modell die Form $\dot{Z} \longrightarrow (\dot{Y}_1, \dot{Y}_2)$ hat, sind $\dot{Y}_1$ und $\dot{Y}_2$ konditional immer unabhängig, denn es gilt dann:

$$\Pr(\dot{Y}_1 = y_1, \dot{Y}_2 = y_2 | \dot{Z} = z) = \Pr(\dot{Y}_1 = y_1 | \dot{Z} = z) \Pr(\dot{Y}_2 = y_2 | \dot{Z} = z)$$

wobei diese Wahrscheinlichkeiten entweder 0 oder 1 sind, je nachdem ob y_1 und y_2 Werte der durch das Modell definierten deterministischen Funktion $f(z)$ sind.

Daraus erkennt man auch, dass wie im deterministischen Fall (vgl. Abschnitt 5.1, § 12) stets eine Zufallsvariable $\dot{U}$ konstruiert werden kann, so dass gilt: $\dot{Y}_1 \perp\!\!\!\perp \dot{Y}_2 | \dot{U}$.[20] Eine solche Variable gehört jedoch meistens nicht zu dem Modell, auf das man sich bezieht.

[19] Als Beispiel kann man wieder an einen Toaster denken, bei dem sowohl die Glühdauer als auch die am Ende der jeweiligen Glühzeit erreichte Temperatur erfasst wird.

[20] Wenn $\dot{Y}_1$ die Werte $a_1, \ldots, a_p$ und $\dot{Y}_2$ die Werte $b_1, \ldots, b_q$ annehmen kann, können pq unterschiedliche Zahlen u_{ij} verwendet und mit einer Wahrscheinlichkeitsverteilung

9. Stochastische Äquivalenz. Die in Abschnitt 5.1 (§ 6) vorgeschlagene Definition für die Äquivalenz funktionaler Modelle ist für stochastische Modelle zwar anwendbar, jedoch selten brauchbar. Oftmals nützlich ist jedoch ein Begriff stochastischer Äquivalenz.

Für die Definition wird der Begriff eines *reduzierten Modells* verwendet. Ist zunächst ein Modell $\mathcal{M} = (\mathcal{V}, \mathcal{F})$ gegeben, wird ein reduziertes Modell in zwei Schritten gebildet: Zuerst werden alle intermediären endogenen Variablen eliminiert; dann werden die stochastischen exogenen Variablen eliminiert, indem die am Ende des ersten Schritts enstandenen Funktionen mit den durch das Modell gegebenen Verteilungen dieser Variablen gemischt (integriert) werden. Ein reduziertes Modell kann also durch eine stochastische Funktion $\ddot{X} \longrightarrow\!\!\!\!\rightarrow \dot{Y}$ charakterisiert werden, wobei $\ddot{X}$ die deterministischen exogenen Modellvariablen und $\dot{Y}$ die nicht-intermediären endogenen Modellvariablen zusammenfasst.[21] Folgende Definition lässt sich anschließen:

– Zwei stochastische Modelle sind *stochastisch äquivalent*, wenn ihre reduzierten Modelle gleich sind.

Insbesondere ist also jedes stochastische Modell mit seiner reduzierten Form stochastisch äquivalent. Als Beispiel kann man an ein Modell der Form $\ddot{X} \longrightarrow \dot{Y} \longleftarrow \dot{Z}$ denken. Das zugehörige reduzierte Modell hat die Form $\ddot{X} \longrightarrow\!\!\!\!\rightarrow \dot{Y}$ (vgl. § 5). Obwohl sich beide Modelle formal und strukturell unterscheiden, sind sie stochastisch äquivalent.

5.3 Einige Fragen der Modellkonzeption

1. Stochastische exogene Variablen? Ein funktionales Modell soll zeigen, wie endogene Variablen (Werte endogener deterministischer bzw. Verteilungen endogener stochastischer Variablen) *von den Werten* exogener Variablen abhängen. Bei den exogenen Variablen kann es sich um determini-

belegt werden. Definiert man dann:

$$\Pr(\dot{Y}_1 = a_i | \dot{U} = u) = \begin{cases} 1 & \text{wenn } u \in \{u_{i1}, \ldots, u_{iq}\} \\ 0 & \text{andernfalls} \end{cases}$$

$$\Pr(\dot{Y}_2 = b_j | \dot{U} = u) = \begin{cases} 1 & \text{wenn } u \in \{u_{1j}, \ldots, u_{pj}\} \\ 0 & \text{andernfalls} \end{cases}$$

$$\Pr(\dot{Y}_1 = a_i, \dot{Y}_2 = b_j | \dot{U} = u) = \begin{cases} 1 & \text{wenn } u = u_{ij} \\ 0 & \text{andernfalls} \end{cases}$$

folgt unmittelbar die konditionale Unabhängigkeit. Man vgl. hierzu auch die Ausführungen von Suppes und Zanotti (1996: 105ff.).

[21] Diese Formulierung gilt zunächst nur für stochastische Modelle i.e.S. Wenn ein stochastisches Modell auch deterministische nicht-intermediäre endogene Variablen enthält, können sie entweder als degenerierte Komponenten in den Vektor $\dot{Y}$ aufgenommen werden, oder man kann zusätzlich eine deterministische Funktion verwenden.

stische und um stochastische Variablen handeln; aber wenn man sich auf Werte einer stochastischen exogenen Variablen bezieht, wird ihre Verteilung gar nicht benötigt und ist irrelevant. Man betrachte als Beispiel die folgenden Modellvarianten:

$$\text{(a)}\quad \begin{array}{l} \ddot{X} \searrow \\ \qquad\quad \dot{Y} \\ \ddot{Z} \nearrow \end{array} \qquad\qquad \text{(b)}\quad \begin{array}{l} \ddot{X} \searrow \\ \qquad\quad \dot{Y} \\ \dot{Z} \nearrow \end{array} \tag{5.4}$$

Der Unterschied besteht darin, dass im Modell (a) eine deterministische Variable $\ddot{Z}$, im Modell (b) eine stochastische Variable $\dot{Z}$ verwendet wird; im Modell (b) muss also zusätzlich eine Annahme über die Verteilung von $\dot{Z}$ getroffen werden. Für die Funktion, die die Modellvariablen verknüpft, ist diese Verteilung jedoch irrelevant, und man kann annehmen, dass die Funktionen in beiden Modellvarianten identisch sind:

$$(x, z) \longrightarrow \Pr[\dot{Y} \mid \ddot{X} = x, \ddot{Z} = z] = \Pr[\dot{Y} \mid \ddot{X} = x, \dot{Z} = z]$$

Wenn man sich auf die Werte exogener Variablen beziehen möchte, genügt es also, sie als deterministische Variablen zu konzipieren. Demgegenüber ist eine Verwendung stochastischer exogener Variablen gerade dann sinnvoll, wenn man keine Annahmen über ihre Werte treffen kann oder treffen möchte. Stattdessen werden Annahmen über ihre Verteilungen getroffen, und diese Annahmen ermöglichen es, die Variablen in einer stochastischen Funktion zu eliminieren. In der Modellvariante (b) kann man eine stochastische Funktion

$$x \longrightarrow \Pr[\dot{Y} | \ddot{X} = x] = \textstyle\sum_{z \in \tilde{\mathcal{Z}}} \Pr[\dot{Y} | \ddot{X} = x, \dot{Z} = z] \Pr(\dot{Z} = z)$$

bilden, in der die Verteilung der endogenen Variablen $\dot{Y}$ nur noch von Werten von $\ddot{X}$ abhängt (vgl. Abschnitt 5.2, § 5).

Infolgedessen stellt sich aber nochmals die Frage, welchem Zweck stochastische exogene Variablen überhaupt dienen können. Eine verbreitete Idee besteht darin, dass solche Variablen zur Repräsentation „unbeobachteter Variablen" verwendet werden können. Was damit gemeint sein kann, hängt allerdings auch davon ab, wie stochastische Modelle konzipiert werden. Besondere Probleme entstehen, wenn mithilfe exogener stochastischer Variablen pseudo-indeterministische Modelle gebildet werden. Da dieser Ansatz sehr verbreitet ist, beginnen wir mit seiner Erläuterung.

2. Pseudo-indeterministische Modelle. Zur Erläuterung kann ein einfaches stochastisches Modell der Form

$$\ddot{X} \longrightarrow\!\!\!\!\rightarrow \dot{Y} \tag{5.5}$$

dienen. Die Überlegung besteht darin, dass es manchmal möglich ist, stattdessen ein stochastisch äquivalentes Modell der Form

$$\ddot{X} \longrightarrow \dot{Y} \longleftarrow \dot{Z} \tag{5.6}$$

zu betrachten, in dem die Modellvariablen durch eine deterministische Funktion verknüpft sind.[22] Infolgedessen verbinden sich mit den beiden Modellvarianten unterschiedliche Interpretationsmöglichkeiten. Die Modellvariante (5.5) erfordert eine stochastische Interpretation: das Modell zeigt, wie *die Verteilung* einer abhängigen Variablen von Werten einer exogenen Variablen abhängt. Dagegen scheint die Modellvariante (5.6) eine deterministische Interpretation zu erlauben; denn sie impliziert die Existenz einer deterministischen Funktion $y = f(x, z)$, durch die *bestimmte Werte* der Variablen $\dot{Y}$ von Werten der exogenen Variablen $\ddot{X}$ und $\dot{Z}$ abhängig sind. Stochastische Modelle der Form (5.6) können deshalb als *pseudo-indeterministische Modelle* bezeichnet worden.[23]

In der Literatur werden funktionale Modelle oftmals von vornherein als pseudo-indeterministische Modelle konzipiert. Den bekanntesten Spezialfall bilden Regressionsmodelle, die in der Form

$$\dot{Y} = g(\ddot{X}) + \dot{Z} \tag{5.7}$$

geschrieben werden, wobei g eine deterministische Funktion ist. Daran knüpfen auch die meisten Autoren an, die sich mit komplexeren funktionalen Modellen beschäftigen.[24] Anders als bei den stochastisch äquivalenten Modellansätzen der Form (5.5) entsteht dann allerdings die Frage, welche Bedeutung den exogenen Zufallsvariablen und ihren Verteilungen gegeben werden kann. Können sie als Repräsentationen „unbeobachteter Variablen" gedeutet werden; und wie kann man Annahmen über ihre Verteilungen verstehen?[25]

3. Definierte unbeobachtete Variablen. Wichtig ist zunächst die Feststellung, dass in zwei wesentlich unterschiedlichen Bedeutungen von „unbeobachteten Variablen" gesprochen werden kann. In einer ersten Bedeutung bezieht sich die Redeweise darauf, dass keine Daten für die Werte einer Variablen verfügbar sind; dabei wird jedoch vorausgesetzt, dass man sich gedanklich auf eine explizit definierte Variable beziehen kann. Wenn diese Bedeutung gemeint ist, sprechen wir von *definierten* unbeobachteten

[22] Ausgehend von einem Modell der Form (5.5) kann natürlich immer auf triviale Weise ein Modell der Form (5.6) konstruiert werden, indem man als Wertebereich für $\dot{Z}$ das kartesische Produkt der Wertebereiche von $\ddot{X}$ und $\dot{Y}$ und zur Definition von $\Pr[\dot{Z}]$ die bedingten Wahrscheinlichkeiten $\Pr(\dot{Y} = y | \ddot{X} = x)$ verwendet.

[23] In einer allgemeinen Bedeutung verwenden wir diese Bezeichnung für stochastische Modelle, bei deren Formulierung nur deterministische Funktionen verwendet werden, so dass sie ihren stochastischen Charakter ausschließlich durch exogene stochastische Variablen gewinnen. In dieser Bedeutung wird die Bezeichnung auch oftmals in der Literatur verwendet, man vgl. beispielsweise Glymour, Spirtes und Scheines (1991: 155-156), Spirtes, Glymour und Scheines (1993: 38-39) und Papineau (2001: 17).

[24] Man vgl. beispielsweise Pearl (2000: 26-27, 44) und Woodward (2003: 339).

[25] Fragen dieser Art werden in der Literatur selten diskutiert, man vgl. jedoch Tryfos (2004: chap. 5).

Variablen. Davon zu unterscheiden sind Redeweisen, durch die die zur Formulierung pseudo-indeterministischer Modelle erforderlichen stochastischen exogenen Variablen als Repräsentationen unbeobachteter Variablen interpretiert werden.

Wir beginnen mit einer Diskussion definierter unbeobachteter Variablen. Solche Variablen können auf unproblematische Weise zur Formulierung funktionaler Modelle verwendet werden. Schwierigkeiten, die aus einem Mangel an Daten resultieren, betreffen nur die Verwendungsmöglichkeiten der Modelle sowie die Möglichkeiten zur Schätzung von Modellparametern mithilfe der jeweils verfügbaren Daten. Zur Verdeutlichung betrachten wir als Beispiel ein Modell

$$\ddot{X} \longrightarrow\!\!\!\!\!\rightarrow \dot{Y} \leftarrow\!\!\!\!\!\longleftarrow \ddot{Z} \tag{5.8}$$

mit dem ein Zusammenhang zwischen Schultyp ($\ddot{X}$) und der Wahrscheinlichkeit, die Schule erfolgreich abzuschließen ($\dot{Y} = 1$, wenn die Schule erfolgreich abgeschlossen wird, andernfalls $\dot{Y} = 0$), untersucht werden soll. Außerdem gibt es eine unbeobachtete Variable $\ddot{Z}$ für das Bildungsniveau der Eltern. Das Modell formuliert also die Hypothese, dass die Wahrscheinlichkeit, die Schule erfolgreich abzuschließen, sowohl vom Schultyp als auch vom Bildungsniveau der Eltern abhängt.

Offenbar ist es für die Möglichkeit, funktionale Modelle formulieren zu können, gleichgültig, ob es sich um unbeobachtete Variablen handelt oder nicht. Es gibt auch zunächst keinen Grund, unbeobachtete Variablen als stochastische Variablen zu konzipieren. Solche Gründe entstehen erst, wenn man das Modell schätzen oder für Prognosen verwenden möchte.

Angenommen, es stehen nur Daten für $\ddot{X}$ und $\dot{Y}$ zur Verfügung. Dann kann das Modell (5.8) nicht geschätzt werden, sondern nur ein Modell der Form (5.5). Daraus entsteht nun auch ein Grund, um anstelle der deterministischen Variablen $\ddot{Z}$ eine Zufallsvariable $\dot{Z}$ zu verwenden: Die Annahme einer Verteilung für $\dot{Z}$ ermöglicht es, das Modell (5.5) als eine reduzierte Form des Modells (5.8) aufzufassen. Natürlich ist das kein Ersatz für die fehlenden Daten; man erreicht jedoch eine konzeptionelle Klärung des Verhältnisses zwischen dem Modell, das man mit den verfügbaren Daten schätzen kann, und dem Modell, auf das sich die theoretischen Hypothesen beziehen. Insbesondere legt das theoretische Modell die Vermutung nahe, dass der durch Daten ermittelbare Zusammenhang zwischen Schultyp und Schulerfolg auch von einer unbekannten Verteilung der Bildungsniveaus der Eltern abhängt. Allerdings muss man bei Formulierungen dieser Art aufpassen. Denn bei der Verteilung der Bildungsniveaus der Eltern handelt es sich um eine statistische Verteilung, zu deren Definition auf eine statistische Gesamtheit Bezug genommen werden muss. Das funktionale Modell, in dem die Zufallsvariable $\dot{Z}$ auftritt, weist jedoch zunächst keinerlei Bezug auf irgendeine statistische Gesamtheit auf; die Verteilung von $\dot{Z}$ kann infolgedessen nicht durch einen gedanklichen Rückgriff auf die Verteilung einer statistischen Variablen erklärt werden.

4. Endogene unbeobachtete Variablen. Unbeobachtete Variablen können auch als endogene Variablen eines stochastischen Modells auftreten. Zur Illustration modifizieren wir das Beispiel des vorangegangenen Paragraphen in folgender Weise:

$$
\begin{tikzcd}
\ddot{Z} \arrow[dd] \arrow[dr] & \\
& \dot{Y} \\
\dot{X} \arrow[ur] &
\end{tikzcd}
\tag{5.9}
$$

Es werden also zwei Hypothesen formuliert: dass die Wahrscheinlichkeit des Schulerfolgs vom Schultyp und vom Bildungsniveau der Eltern abhängt, und dass die Wahl eines Schultyps ($\dot{X}$) durch eine stochastische Funktion vom Bildungsniveau der Eltern abhängt.

Für das Beispiel soll nun angenommen werden, dass $\dot{X}$ eine unbeobachtete Variable ist, also Daten nur für das Bildungsniveau der Eltern ($\ddot{Z}$) und den Schulerfolg ($\dot{Y}$) verfügbar sind. Mit diesen Daten kann also wiederum nur ein reduziertes Modell $\ddot{Z} \longrightarrow\!\!\!\!\rightarrow \dot{Y}$ mit der stochastischen Funktion

$$z \longrightarrow \Pr[\dot{Y}|\ddot{Z}=z] = \sum_x \Pr[\dot{Y}|\ddot{Z}=z,\dot{X}=x]\,\Pr(\dot{X}=x|\ddot{Z}=z)$$

geschätzt werden. Anders als bei einer exogenen unbeobachteten Variablen ist jetzt jedoch keine separate Verteilungsannahme erforderlich. Ein reduziertes Modell kann ohne weitere Annahmen aus dem zunächst konzipierten theoretischen Modell abgeleitet werden. Zwar hängt die stochastische Funktion für das reduzierte Modell von den bedingten Verteilungen $\Pr[\dot{X}|\ddot{Z}=z]$ ab, die mit den verfügbaren Daten nicht geschätzt werden können; wird das Modell (5.9) vorausgesetzt, ist es jedoch nicht erforderlich (und auch gar nicht möglich), Annahmen über die Verteilung der gewählten Schultypen zu treffen.

5. Annahmen über Verteilungen. Es gibt also einen wesentlichen Unterschied zwischen exogenen und endogenen stochastischen Variablen. Bei exogenen Zufallsvariablen müssen Annahmen über ihre Verteilungen gemacht werden; dagegen werden über Verteilungen endogener stochastischer Variablen weder direkt noch indirekt Annahmen getroffen. Vielmehr werden stochastische Funktionen spezifiziert, die zeigen, wie die Verteilungen endogener stochastischer Variablen von Werten anderer Modellvariablen abhängen. Infolgedessen gibt es für diese Variablen ausschließlich konditionale Verteilungen.

Man denke zum Beispiel an die Variable $\dot{X}$ im Modell (5.9), die die Wahl eines Schultyps erfasst. Es handelt sich zwar um eine stochastische Variable, gleichwohl kann man ihr innerhalb des Modells keine bestimmte Verteilung zurechnen. Denn das Modell zeigt nur, wie ihre Verteilung von Werten einer exogenen Variablen $\ddot{Z}$ abhängt, und über deren Werte werden

im Modell keine Annahmen getroffen. Es gibt auch keinen unmittelbaren Zusammenhang zwischen der Zufallsvariablen $\dot{X}$ und einer statistischen Variablen X, durch die erfasst wird, wie sich in irgendeiner statistischen Gesamtheit Schüler auf Schultypen verteilen. Ein funktionales Modell bezieht sich zunächst überhaupt nicht auf statistische Gesamtheiten, und sowohl deterministische als auch stochastische Modellvariablen müssen von statistischen Variablen grundsätzlich unterschieden werden.

Anders als bei endogenen Variablen sind bei exogenen stochastischen Variablen Annahmen über ihre Verteilung erforderlich. Somit stellt sich auch die Frage, welche Bedeutung diesen Verteilungsannahmen zukommt und von welchen Überlegungen man sich bei ihrer Formulierung leiten lassen kann. Offenbar muss zunächst überlegt werden, warum es überhaupt erforderlich erscheint, exogene Variablen als stochastische Variablen zu konzipieren. Denn da mithilfe funktionaler Modelle untersucht werden soll, wie endogene Variablen von den Werten exogener Variablen abhängen, sollte versucht werden, alle exogenen Variablen deterministisch zu konzipieren. Stochastische exogene Variablen sind nur erforderlich, wenn keine (oder nur ungenaue[26]) Daten verfügbar sind und (deshalb) ein reduziertes Modell gebildet werden soll.

Weitere Überlegungen hängen davon ab, ob sich das Problem fehlender Daten bei der Schätzung oder bei der Verwendung eines Modells stellt. Im ersten Fall sind Annahmen über Verteilungen unbeobachteter Variablen (oder auch nur über einige ihrer Eigenschaften) oft schon erforderlich, um (parametrische) Modelle formulieren zu können, die mit den verfügbaren Daten geschätzt werden können. (Das wird in Kapitel 7 etwas ausführlicher besprochen.)

Anders verhält es sich, wenn sich das Problem bei der Verwendung eines bereits verfügbaren vollständigen Modells stellt. Als Beispiel kann das Modell (5.8) dienen. Angenommen, es soll verwendet werden, um bei irgendeinem bestimmten Kind (oder bei irgendeiner bestimmten Gesamtheit von Kindern) Vermutungen darüber anzustellen, wie der Schulerfolg vom Schultyp abhängt. Hätte man Informationen über das Bildungsniveau der Eltern des Kindes, könnte das Modell ohne weiteres verwendet werden. Fehlt diese Information, kann man sich entweder auf konditionale Vermutungen beschränken; oder man kann eine subjektive Wahrscheinlichkeitsverteilung für die unbeobachtete Variable annehmen, so dass man subjektive Erwartungswerte formulieren kann.

6. Fiktive Residualvariablen. Es sollte deutlich geworden sein, dass definierte unbeobachtete Variablen konzeptionell unproblematisch sind. Dagegen treten bei stochastischen Variablen, die zur Formulierung pseudoindeterministischer Modelle verwendet werden, von vornherein konzeptionelle Probleme auf. Man betrachte als Beispiel das Modell (5.6). Um die

[26] Hier kommen natürlich nur Meßfehler in Betracht, die im Kontext der intendierten Modellverwendung relevant sind.

Verwendung einer deterministischen Funktion $y = f(x, z)$ zu rechtfertigen, muss angenommen werden, dass die stochastische Variable $\dot{Z}$ *alle* Bedingungen (außer $\ddot{X}$) erfasst, von denen Werte von $\dot{Y}$ abhängig sind. Dieser Variablen kann infolgedessen keine bestimmte Bedeutung gegeben werden.

Um die Idee zu verfolgen, dass sich *alle* Bedingungen erfassen lassen, von denen $\dot{Y}$ abhängt, müsste man zunächst von einem Modell der Form

$$\ddot{X} \longrightarrow \dot{Y} \longleftarrow (\dot{Z}_1, \ldots, \dot{Z}_m) \tag{5.10}$$

ausgehen, wobei die Variablen $\dot{Z}_1, \ldots, \dot{Z}_m$ die unterschiedlichen Bedingungen erfassen. Aber offenbar ist es unmöglich, ein Modell dieser Art zu konzipieren, da niemand eine vollständige Liste dieser Variablen angeben könnte. Aber selbst wenn es möglich wäre, eine solche Liste vorauszusetzen, bliebe immer noch die Frage, wie eine Funktion

$$\phi : \tilde{\mathcal{Z}}_1 \times \cdots \times \tilde{\mathcal{Z}}_m \longrightarrow \tilde{\mathcal{Z}}$$

definiert werden könnte, die die Variablen $\dot{Z}_1, \ldots, \dot{Z}_m$ zu einer Variablen $\dot{Z}$ zusammenfasst, und wie die Annahme einer Verteilung für $\dot{Z}$ begründet werden könnte.

Fragen dieser Art werden indessen bei der Verwendung pseudo-indeterministischer Modelle gar nicht überlegt. Stattdessen wird $\dot{Z}$ als eine *fiktive Residualvariable* konzipiert, die dem Zweck dient, ein pseudo-indeterministisches Modell formulieren zu können, und die deshalb als eine unbeobachtete Variable interpretiert wird, die alle nicht explizit durch Modellvariablen erfassten Bedingungen zusammenfasst. Wie bei definierten unbeobachteten Variablen gibt es dann aber zunächst gar keinen Grund, von einer Zufallsvariablen $\dot{Z}$ auszugehen; vielmehr wäre von einer deterministischen Variablen $\ddot{Z}$, also von einem Modell der Form

$$\ddot{X} \longrightarrow \ddot{Y} \longleftarrow \ddot{Z} \tag{5.11}$$

auszugehen. Eine Begründung dafür, anstelle von $\ddot{Z}$ eine Zufallsvariable $\dot{Z}$ anzunehmen, entsteht erst durch die Absicht, anstelle von (5.11) ein reduziertes Modell zu verwenden. Dieses reduzierte Modell hat jedoch nicht die Form (5.6), sondern die Form (5.5), also eine Form, in der die fiktive Residualvariable nicht mehr erscheint.

Diese Überlegung erlaubt auch eine Klärung der gelegentlich diskutierten Frage, ob man bei einem pseudo-indeterministischen Modell annehmen kann, dass die fiktive Residualvariable von den deterministischen exogenen Variablen stochastisch unabhängig ist; in unserem Beispiel: ob $\dot{Z}$ von $\ddot{X}$ stochastisch unabhängig ist.[27] Da die fiktive Residualvariable nur dem Zweck dient, ein reduziertes Modell bilden zu können, ist die Frage eigentlich gegenstandslos; denn innerhalb des reduzierten Modells kann sie

[27] Man vgl. beispielsweise Freedman (1992), Clogg und Haritou (1997), Woodward (2003: 325-327).

gar nicht gestellt werden.[28] Gleichwohl kann man natürlich ausgehend von dem Modell (5.6) über die Möglichkeit nachdenken, dass $\dot{Z}$ von $\ddot{X}$ stochastisch abhängig ist. Dann entsteht jedoch ein neues Modell, das folgende Form hat:

$$\begin{tikzcd}
\ddot{X} \arrow[dr] \arrow[dd] & \\
& \dot{Y} \\
\dot{Z} \arrow[ur] &
\end{tikzcd} \tag{5.12}$$

Offenbar wird dann $\dot{Z}$ zu einer endogenen Variablen, und es handelt sich nicht mehr um ein pseudo-indeterministisches Modell. Vielmehr entsteht ein Modell, das nur noch eine nicht-stochastische exogene Variable verwendet. Die Vermutung, dass eine fiktive Residualvariable von explizit definierten exogenen Variablen stochastisch abhängig sein könnte, impliziert also die Vermutung, dass ein pseudo-indeterministisches Modell nicht angemessen ist.[29]

Einer solchen Vermutung kann indessen leicht Rechnung getragen werden, indem man zur Modellformulierung deterministische exogene Variablen und stochastische Funktionen verwendet. Offenbar gelangt man sowohl ausgehend von (5.6) als auch ausgehend von (5.12) zu einem reduzierten Modell der Form (5.5).[30]

[28] Anders verhält es sich jedoch bei definierten unbeobachteten Variablen, bei denen Unabhängigkeitsfragen insbesondere für Möglichkeiten kausaler Modellinterpretationen relevant sind. Das wird in Kapitel 6 genauer besprochen.

[29] Die Annahme ist also für die Formulierung pseudo-indeterministischer Modelle konstitutiv. Diese Auffassung scheinen auch Pratt und Schlaifer (1984) zu vertreten.

[30] Geht man von (5.12) aus, erhält man für das reduzierte Modell anstelle von (5.2) die stochastische Funktion $x \longrightarrow \Pr(\dot{Y} = y|\ddot{X} = x) = \sum_{z:f(x,z)=y} \Pr(\dot{Z} = z|\ddot{X} = x)$. Der Unterschied betrifft also nur die Darstellungen der bedingten Verteilungen von $\dot{Y}$ als fiktive Mischungen.

5.4 Indirekte Prognosefunktionen

1. Direkte und indirekte Prognosen. Zur Erläuterung der Unterscheidung beziehen wir uns zunächst auf ein deterministisches Modell $\mathcal{M} = (\mathcal{V}, \mathcal{F})$. Alle definierten Funktionen (also die Elemente von $\mathcal{F}$) sowie alle Funktionen, die durch eine Verkettung definierter Funktionen gebildet werden können, werden als *direkte Prognosefunktionen* bezeichnet. So gibt es zum Beispiel in dem Modell

$$\ddot{X} \longrightarrow \ddot{Y} \longleftarrow \ddot{Z} \longleftarrow \ddot{V} \tag{5.13}$$

drei direkte Prognosefunktionen: zwei definierte Funktionen und eine abgeleitete Funktion mit den Argumentvariablen $\ddot{X}$ und $\ddot{V}$ und der abhängigen Variablen $\ddot{Y}$. Funktionen, die für eindeutige oder unbestimmte Prognosen verwendet, aber nicht durch eine Verkettung definierter Funktionen gebildet werden können, bezeichnen wir als *indirekte Prognosefunktionen*. Wie in Abschnitt 5.1 besprochen wurde, können sie manchmal als gewöhnliche Funktionen, jedoch immer als mengenwertige Funktionen gebildet werden. Zum Beispiel könnte bei dem Modell (5.13) eine Funktion

$$v \longrightarrow \mathrm{B}[\ddot{X}|\ddot{V}=v] \tag{5.14}$$

gebildet werden, die die Variablen $\ddot{V}$ und $\ddot{X}$ verknüpft; und es wäre eine indirekte Prognosefunktion, wenn für mindestens einen Wert v die Menge $\mathrm{B}[\ddot{X}|\ddot{V}=v]$ der durch v bedingten möglichen Werte für $\ddot{X}$ weniger Elemente enthält als die Menge $\mathrm{B}[\ddot{X}]$ der unbedingt möglichen Werte.

Eine analoge Unterscheidung zwischen direkten und indirekten Prognosefunktionen kann für stochastische Modelle vorgenommen werden. In Analogie zu (5.13) beziehen wir uns als Beispiel auf ein Modell

$$\ddot{X} \twoheadrightarrow \dot{Y} \twoheadleftarrow \dot{Z} \twoheadleftarrow \ddot{V} \tag{5.15}$$

Wiederum gibt es drei direkte Prognosefunktionen, bei denen es sich jetzt um stochastische Funktionen handelt. Zu überlegen ist, ob bzw. wie auch indirekte Prognosefunktionen gebildet werden können. Zwar können auch bei stochastischen Modellen mengenwertige Funktionen konstruiert werden; es ist jedoch selten der Fall, dass sie für Prognosen verwendet werden können. Eine andere Möglichkeit, mit der wir uns im Folgenden beschäftigen, besteht darin, neue stochastische Funktionen zu konstruieren.

2. Umkehrungen stochastischer Funktionen. Allerdings ist das ohne weiteres nicht möglich. Als Beispiel betrachten wir eine stochastische Funktion

$$\dot{X} \twoheadrightarrow \dot{Y} \tag{5.16}$$

Ohne zusätzliche Annahmen ist es nicht möglich, daraus eine stochastische Funktion $\dot{Y} \twoheadrightarrow \dot{X}$ abzuleiten, die es erlauben würde, Werte von $\dot{Y}$ für

Prognosen über die Verteilung von $\dot{X}$ zu verwenden. Erforderlich ist eine Annahme über $\Pr[\dot{X}]$, also die unbedingte Verteilung von $\dot{X}$. Dann kann zunächst die gemeinsame Verteilung

$$\Pr(\dot{X}=x, \dot{Y}=y) = \Pr(\dot{Y}=y|\dot{X}=x)\,\Pr(\dot{X}=x)$$

gebildet werden, die dann die stochastische Funktion $\dot{Y} \twoheadrightarrow \dot{X}$ liefert.

Die entscheidende Frage ist offenbar, wie Annahmen über unbedingte Verteilungen gewonnen und begründet werden können. Denn abgesehen von dem problematischen Sonderfall abgeschlossener stochastischer Modelle können solche Annahmen nicht aus dem Modell abgeleitet werden.

3. Überlegungen anhand eines Beispiels. Um das Problem zu verdeutlichen, beziehen wir uns auf ein einfaches Beispiel: $\ddot{X} \twoheadrightarrow \dot{Y}$. $\ddot{X}$ erfasst, ob ein bestimmtes Medikament eingenommen wurde ($\ddot{X}=1$) oder nicht ($\ddot{X}=0$), und $\dot{Y}$ erfasst, ob ein Heilungserfolg eingetreten ist ($\dot{Y}=1$) oder nicht ($\dot{Y}=0$). Die stochastische Funktion sei durch

$$\Pr(\dot{Y}=1|\ddot{X}=0) = 0.2 \quad \text{und} \quad \Pr(\dot{Y}=1|\ddot{X}=1) = 0.7$$

gegeben. Das indirekte Prognoseproblem soll darin bestehen, aufgrund der Kenntnis, dass ein Heilungserfolg eingetreten ist, eine Vermutung darüber anzustellen, ob das Medikament eingenommen wurde.

Offenbar ist weder eine eindeutige noch (mittels einer mengenwertigen Funktion) eine unbestimmte Prognose möglich. Ein probabilistischer Ansatz beginnt stattdessen mit der Idee, die deterministische Variable $\ddot{X}$ durch eine stochastische Variable $\dot{X}$ zu ersetzen, so dass das Prognoseproblem reformuliert werden kann. Es besteht dann darin, ausgehend von einem bestimmten Wert von $\dot{Y}$ eine Vermutung über die Verteilung von $\dot{X}$ zu begründen.

Als konkrete Problemvariante nehmen wir an, dass man weiß, dass ein Heilungserfolg eingetreten ist: $\dot{Y}=1$. Aber diese Voraussetzung genügt offenbar nicht, um eine durch $\dot{Y}=1$ bedingte Verteilung von $\dot{X}$ zu bilden. Man benötigt zunächst eine unbedingte Verteilung; zum Beispiel könnte man annehmen, dass beide Möglichkeiten gleichwahrscheinlich sind: $\Pr(\dot{X}=1)=0.5$. Dann kann man die gemeinsame Verteilung berechnen:

$$\Pr(\dot{X}=x, \dot{Y}=y) = \begin{cases} 0.8 \cdot 0.5 = 0.40 & \text{wenn } x=0 \text{ und } y=0 \\ 0.2 \cdot 0.5 = 0.10 & \text{wenn } x=0 \text{ und } y=1 \\ 0.3 \cdot 0.5 = 0.15 & \text{wenn } x=1 \text{ und } y=0 \\ 0.7 \cdot 0.5 = 0.35 & \text{wenn } x=1 \text{ und } y=1 \end{cases}$$

und findet als Antwort für das Prognoseproblem:

$$\Pr(\dot{X}=1|\dot{Y}=1) = 0.35/0.45 \approx 0.78$$

Aber diese Antwort ist offenbar wesentlich von der anfänglichen Annahme über die unbedingte Verteilung von $\dot{X}$ abhängig. Mit anderen Annahmen

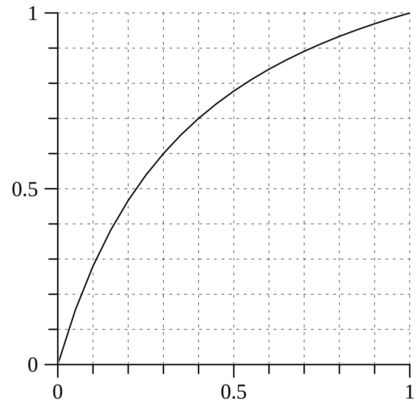


Abb. 5.4-1 Abhängigkeit von $\Pr(\dot{X} = 1|\dot{Y} = 1)$ (Y-Achse) von Annahmen über $\Pr(\dot{X} = 1)$ (X-Achse).

würde man zu anderen Prognosen gelangen; Abbildung 5.4-1 zeigt das für unser Beispiel.

Kapitel 6

Funktionale Kausalität

In diesem Kapitel wird die Diskussion funktionaler Modelle fortgesetzt. Im Mittelpunkt steht die Frage, wie man bezugnehmend auf solche Modelle von kausalen Zusammenhängen sprechen kann. Die leitende Idee besteht darin, Werte von Variablen, die *Bedingungen für Werte* anderer Variablen sein können, und Veränderungen der Werte von Variablen, die *Ursachen für Veränderungen* der Werte anderer Variablen sein können, zu unterscheiden. Da sich die Definitionen auf funktionale Modelle beziehen, sprechen wir von funktionaler Kausalität. Es gibt andere Möglichkeiten, von kausalen Zusammenhängen zu sprechen, mit denen wir uns hier nicht näher beschäftigen; und auch funktionale Kausalität liefert nur einen formalen Rahmen, der mit unterschiedlichen Konzeptionen kausaler Zusammenhänge vereinbar ist.

Es gibt drei Abschnitte. Im ersten Abschnitt werden Definitionen funktionaler Kausalität für deterministische und stochastische Modelle besprochen. Dabei wird betont, dass sich die Definitionen auf Modelle und nicht unmittelbar auf die Realität beziehen. Im zweiten Abschnitt beschäftigen wir uns mit einer in der statistischen Literatur verbreiteten Idee: mithilfe kontrafaktischer Variablen kausale Effekte für identifizierbare Individuen zu definieren. Es wird gezeigt, wie man einige problematische Aspekte dieses Ansatzes vermeiden kann, wenn man auf die Annahme identifizierbarer Individuen verzichtet und stattdessen eine statistische Betrachtungsweise einnimmt. Der dritte Abschnitt verfolgt die Idee, dass es möglich sein sollte, Ursachen bestimmte isolierbare Wirkungen zuzurechnen. Es wird gezeigt, dass das nicht immer möglich ist und dass sich daraus Sinngrenzen für Feststellungen über funktionale Kausalzusammenhänge ergeben können.

6.1 Bedingungen und Veränderungen

In diesem Abschnitt werden Definitionen funktionaler Kausalität besprochen. In den Notationen wird an die Ausführungen über funktionale Modelle in Kapitel 5 angeknüpft.

1. Deterministische Modelle. Wir beziehen uns auf ein deterministisches funktionales Modell $\mathcal{M} = (\mathcal{V}, \mathcal{F})$ und nehmen an, dass $\ddot{X}$ und $\ddot{Y}$ zwei Modellvariablen aus $\mathcal{V}$ sind. Definiert werden soll, was damit gemeint ist, dass eine Veränderung $\Delta(x', x'')$ bei der Variablen $\ddot{X}$ eine *funktionale Ursache* einer Veränderung $\Delta(y', y'')$ bei der Variablen $\ddot{Y}$ ist.[1]

Als Hilfsmittel wird folgende Sprechweise verwendet: Ein *Kovariablenkontext* einer Variablen $\ddot{X}$ mit der Belegung $\ddot{X} = x$ bezüglich einer von $\ddot{X}$ funktional abhängigen Variablen $\ddot{Y}$ besteht aus einer Variablen $\ddot{Z} \in \mathcal{V}$ und einer Belegung $\ddot{Z} = z$ dieser Variablen,[2] so dass aus einer Kenntnis dieser Belegung in Verbindung mit $\ddot{X} = x$ eindeutig ein Wert von $\ddot{Y}$ berechnet werden kann. Natürlich kann es sein, dass kein Kovariablenkontext erforderlich ist, um bei einer Kenntnis von $\ddot{X} = x$ einen Wert von $\ddot{Y}$ zu berechnen, und es ist auch klar, dass es im Allgemeinen unterschiedliche Kovariablenkontexte geben kann.

Ausgehend von dieser Sprechweise können folgende Definitionen funktionaler Kausalität gegeben werden:

– Eine Veränderung $\Delta(x', x'')$ bei der Variablen $\ddot{X}$ ist im Kovariablenkontext $\ddot{Z} = z$ eine *funktionale Ursache* einer Veränderung $\Delta(y', y'')$ bei der Variablen $\ddot{Y}$, wenn gilt: $\ddot{Y}$ ist von $\ddot{X}$ funktional abhängig und bei gleichbleibendem Kovariablenkontext $\ddot{Z} = z$ ist $\ddot{Y} = y'$ aus $\ddot{X} = x'$ und $\ddot{Y} = y''$ aus $\ddot{X} = x''$ berechenbar.

– Eine Veränderung $\Delta(x', x'')$ bei der Variablen $\ddot{X}$ ist eine *kontextunabhängige funktionale Ursache* einer Veränderung $\Delta(y', y'')$ bei der Variablen $\ddot{Y}$, wenn kein Kovariablenkontext erforderlich ist oder bei allen im Modell möglichen Kovariablenkontexten $\ddot{Y} = y'$ aus $\ddot{X} = x'$ und $\ddot{Y} = y''$ aus $\ddot{X} = x''$ berechnet werden kann.

[1] Mit der Notation $\Delta(x', x'')$ ist gemeint, dass sich der Wert der betreffenden Variablen von x' zu x'' ändert. Die Notation impliziert $x' \neq x''$.

[2] Anstelle von $\ddot{Z}$ kann auch ein Vektor $(\ddot{Z}_1, \ldots, \ddot{Z}_m)$ verwendet werden, dessen Komponenten Elemente von $\mathcal{V}$ sind.

In komplementärer Weise kann man Veränderungen $\Delta(y', y'')$ als *funktionale Wirkungen* oder *Effekte* ihrer (kontextabhängigen oder -unabhängigen) funktionalen Ursachen bezeichnen. – Folgende Aspekte dieser Definitionen sind bemerkenswert:

a) Beim Reden von funktionaler Kausalität wird stets auf ein funktionales Modell Bezug genommen, das zunächst angegeben werden muss. Mit diesem Kausalitätsbegriff können also nicht unmittelbar empirische Aussagen formuliert werden. Dies ist auch wichtig, um zu verstehen, was mit Veränderungen der Werte von Variablen gemeint ist: Solche Veränderungen, wie sie etwa durch $\Delta(x', x'')$ vergegenwärtigt werden, entstehen im Rahmen eines Modells durch Annahmen bzw. Setzungen des Modellkonstrukteurs. (Zunächst wird auch offen gelassen, unter welchen Bedingungen solche Annahmen über Veränderungen der Werte von Modellvariablen sinnvoll getroffen werden können.)

b) Funktionale Kausalität verknüpft *Veränderungen* der Werte von Variablen.[3] Orientiert man sich an den vorgeschlagenen Definitionen, können weder Variablen noch bestimmte Werte von Variablen als funktionale Ursachen anderer Variablen oder ihrer Werte bezeichnet werden.[4] Bei Werten von Variablen kann man jedoch unter Umständen davon sprechen, dass sie *funktionale Bedingungen* für bestimmte Werte anderer Variablen sind.[5] So kann man in dem Glühlampenbeispiel aus Abschnitt 5.1 (§ 1) sagen, dass das Geschlossensein des Schalters ($\ddot{X} = 1$) eine funktionale Bedingung für das Leuchten der Glühbirne ($\ddot{Y} = 1$) ist. Dagegen kann nur eine Veränderung wie beispielsweise das Schließen des Schalters ($\Delta(0, 1)$ bei der Variablen $\ddot{X}$) unter Umständen (im Kovariablenkontext $\ddot{Z} = 1$, d.h. wenn die Batterie Strom liefert) eine funktionale Ursache einer anderen Veränderung ($\Delta(0, 1)$ bei der Variablen $\ddot{Y}$) sein.[6]

[3] Dies entspricht der oft geäußerten Idee, dass kausale Redeweisen auf Beziehungen zwischen Ereignissen bzw. Ereignistypen verweisen sollen; man vgl. etwa Hausman (1998: chap. 2).

[4] Wir vermeiden deshalb nicht nur Formulierungen, in denen bestimmte Werte von Variablen als Ereignisse bezeichnet werden (wie beispielsweise bei Granger 1990: 45), sondern auch Formulierungen, in denen unspezifisch von „(kausalen) Faktoren" gesprochen wird, ohne deutlich zwischen (funktionalen) Bedingungen und Ursachen zu unterscheiden. Natürlich können abkürzende Redeweisen verwendet werden, wenn ihre Bedeutung aus dem Kontext ersichtlich ist; man vgl. auch die Bemerkungen von Woodward (2003: 39).

[5] Hier knüpft ein wesentlich anderer Sprachgebrauch an, der sich insbesondere seit John St. Mill verbreitet hat: nämlich als Ursachen eines Sachverhalts dessen (notwendige oder hinreichende) Bedingungen zu bezeichnen; man vgl. beispielsweise Rothman und Greenland (2005) und Susser (1991). Es ist wichtig, diesen Sprachgebrauch nicht mit dem hier vorgeschlagenen Reden von funktionaler Kausalität zu verwechseln.

[6] Ein anderes Beispiel, an dem die Bedeutung der Unterscheidung gut sichtbar wird, findet man bei Hausman (1998: 25).

c) Die Idee, funktionale Ursachen *als Veränderungen* von (funktionalen) *Bedingungen* zu unterscheiden, ist für den hier verfolgten Ansatz von grundlegender Bedeutung, zugleich jedoch eine Quelle möglicher Unklarheiten. Denn wenn nur abstrakt auf ein funktionales Modell Bezug genommen wird, kann nicht in (empirisch) gehaltvoller Weise von Veränderungen gesprochen werden, insbesondere steht kein gehaltvoller Ereignisbegriff zur Verfügung, sondern es können nur unterschiedliche (funktionale) Bedingungen verglichen werden.[7]

d) Funktionale Kausalität ist in den meisten Fällen kontextabhängig, d.h. ob und wie eine Veränderung $\Delta(x', x'')$ funktionale Ursache einer Veränderung $\Delta(y', y'')$ ist, hängt meistens auch von den Werten anderer Variablen ab. Zum Beispiel ist das Schließen des Schalters nur dann eine funktionale Ursache dafür, dass die Glühbirne zu leuchten anfängt, wenn die Batterie Strom liefert. Diese Kontextabhängigkeit entspricht dem in Abschnitt 5.1 (§ 4) betonten Umstand, dass die in einem funktionalen Modell relevanten Abhängigkeiten zunächst durch Funktionen festgestellt werden; dagegen kann den Pfeilen, durch die einzelne Variablen verknüpft werden, im Allgemeinen keine bestimmte kausale Bedeutung gegeben werden.

e) Eine Veränderung $\Delta(x', x'')$ des Werts einer Variablen $\ddot{X}$ kann nur dann eine funktionale Ursache der Veränderung des Werts einer anderen Variablen $\ddot{Y}$ sein, wenn die Veränderung $\Delta(x', x'')$ möglich ist, ohne dass sich zugleich Werte anderer Variablen, von denen $\ddot{Y}$ auch abhängt, verändern. Dies kann unter Umständen, wenn es Constraints gibt oder bei endogenen Variablen, nicht oder nur eingeschränkt möglich sein. Eine weitere Restriktion resultiert aus der Kontextabhängigkeit: Damit $\Delta(x', x'')$ eine funktionale Ursache sein kann, muss bei beiden Werten von $\ddot{X}$ *der gleiche* Kovariablenkontext möglich sein. Offenbar kann es dafür Restriktionen geben, wenn einige Kovariablen selbst von $\ddot{X}$ abhängig sind. In solchen Fällen kann $\Delta(x', x'')$ kein isolierbarer Effekt zugerechnet werden. (Mit Problemen dieser Art beschäftigen wir uns in Abschnitt 6.3.)

f) Da funktionale Modelle nicht unbedingt eine zeitliche Ordnung ihrer Variablen voraussetzen, wird auch bei der Definition funktionaler Kausalität keine zeitliche Ordnung gefordert. Natürlich ist es möglich, bei dynamischen funktionalen Modellen, die sich als temporale Ablaufschemas interpretieren lassen, zeitliche Ordnung als eine zusätzliche Bedingung einzuführen.

Die grundlegende Idee besteht also in einer Unterscheidung zwischen Ursachen und Bedingungen. Diese Unterscheidung ermöglicht auch ein besseres Verständnis der oft geäußerten Behauptung, dass am Zustandekommen

[7] Mit einigen daraus resultierenden Problemen und mit alternativen konzeptionellen Ansätzen beschäftigen wir uns in Kapitel 8.

von Ereignissen oft mehrere, viele oder sogar sehr viele Ursachen beteiligt sind.[8] Denn man kann sich zwar meistens beliebig viele Bedingungen vorstellen, von denen das Eintreten eines Ereignisses abhängig sein kann; das Reden von Ursachen zielt aber normalerweise auf eine bestimmte Ursache (Veränderung), die *unter gegebenen Bedingungen* das Ereignis zur Folge hat. Deshalb ist es für den Ursachenbegriff wesentlich, dass Kontextbedingungen konstantgehalten werden können.

2. Definitionen für stochastische Modelle. Zu überlegen ist, wie auch für stochastische Modelle eine Definition funktionaler Kausalität gegeben werden kann. Bezieht man sich auf zwei stochastische Variablen, die durch eine deterministische Funktion verknüpft sind, etwa $\dot{X} \longrightarrow \dot{Y}$ (ggf. unter Beteiligung weiterer Kovariablen), können die Definitionen aus § 1 übernommen werden, also Veränderungen $\Delta(x', x'')$ der Werte von $\dot{X}$ als mögliche, ggf. kontextabhängige funktionale Ursachen von Veränderungen der Werte von $\dot{Y}$ betrachtet werden. Wir beziehen uns deshalb im Folgenden auf eine endogene stochastische Variable $\dot{Y}$, deren Werte durch eine stochastische Funktion von Werten anderer Modellvariablen abhängig sind.

Betrachten wir zunächst ein einfaches Modell der Form $\ddot{X} \longrightarrow\!\!\!\!\rightarrow \dot{Y}$, zum Beispiel (wie in Abschnitt 5.2, § 2) einen Toaster, so dass sich $\ddot{X}$ auf die Einstellung eines Hebels und $\dot{Y}$ auf die Glühdauer bezieht. Zu einer Veränderung $\Delta(x', x'')$ erhält man dann nicht mehr bestimmte Werte von $\dot{Y}$, sondern zwei bedingte Verteilungen: $\Pr[\dot{Y}|\ddot{X} = x']$ und $\Pr[\dot{Y}|\ddot{X} = x'']$. Infolgedessen gibt es unterschiedliche Möglichkeiten, um einer Veränderung $\Delta(x', x'')$ zurechenbare kausale Effekte zu definieren. Da es in dem Beispiel keinen Kovariablenkontext gibt, könnte man die Differenz der Erwartungswerte

$$\mathrm{E}(\dot{Y}|\ddot{X} = x') - \mathrm{E}(\dot{Y}|\ddot{X} = x)$$

verwenden. Da auch andere Definitionen möglich sind, sprechen wir im Folgenden allgemein von einem *stochastischen Effekt* und verwenden Bezeichnungen der Form $\Delta^s(\dot{Y}; x', x'')$.

Übernimmt man nun sinngemäß das Reden von Kovariablenkontexten aus § 1, können folgende Definitionen verwendet werden; dabei wird angenommen, dass $\ddot{X}$ und $\dot{Y}$ Variablen eines stochastischen funktionalen Modells $\mathcal{M} = (\mathcal{V}, \mathcal{F})$ sind.

– Eine Veränderung $\Delta(x', x'')$ bei der Variablen $\ddot{X}$ ist im Kovariablenkontext $\ddot{Z} = z$ eine *funktionale Ursache* für einen stochastischen Effekt $\Delta^s(\dot{Y}; x', x'')$, wenn gilt: Es gibt einen gerichteten Pfad von $\ddot{X}$ zu $\dot{Y}$, und der stochastische Effekt ist aus den bedingten Verteilungen $\Pr[\dot{Y}|\ddot{X} = x', \ddot{Z} = z]$ und $\Pr[\dot{Y}|\ddot{X} = x'', \ddot{Z} = z]$ berechenbar.

[8] Man vgl. beispielsweise Marini und Singer (1988: 354). Urheber der Idee ist John St. Mill, der ausdrücklich keine Unterscheidung zwischen Ursachen und Bedingungen vorgenommen hat.

– Eine Veränderung $\Delta(x', x'')$ bei der Variablen $\ddot{X}$ ist eine *kontextunabhängige funktionale Ursache* eines stochastischen Effekts $\Delta^s(\dot{Y}; x', x'')$, wenn gilt: Es gibt einen gerichteten Pfad von $\ddot{X}$ zu $\dot{Y}$, und der stochastische Effekt ist aus den bedingten Verteilungen $\Pr[\dot{Y}|\ddot{X} = x']$ und $\Pr[\dot{Y}|\ddot{X} = x'']$ berechenbar und hängt nicht von Werten anderer Modellvariablen ab.

Bei diesen Definitionen wird angenommen, dass die Verteilung von $\dot{Y}$ nur von deterministischen Modellvariablen abhängt. Die Definitionen können jedoch sinngemäß auch verwendet werden, wenn die beteiligten Modellvariablen stochastisch sind ($\dot{X}$ und/oder $\dot{Z}$).[9]

Die Erläuterungen (a) – (f) aus § 1 gelten sinngemäß für stochastische Effekte. Wiederum ist zu betonen, dass von *Veränderungen* $\Delta(x', x'')$ ausgegangen wird.[10] Außerdem ist wichtig, dass auch bei stochastischen Effekten funktionale Ursachen im Allgemeinen kontextabhängig sind. Beispielsweise ist es möglich, dass eine Veränderung $\Delta(x', x'')$ die Wahrscheinlichkeit eines Ereignistyps ($\Pr(\dot{Y} = 1)$ bei einer binären Variablen $\dot{Y}$) je nach Beschaffenheit eines Kovariablenkontextes sowohl erhöhen als auch verringern kann. Es erscheint deshalb nicht sinnvoll, von vornherein kontextunabhängige Definitionen zu verlangen.[11] Insbesondere erscheint es nicht sinnvoll, bereits qua Definition zu fordern, dass eine Ursache die Wahrscheinlichkeit einer bestimmten möglichen Wirkung stets (in allen möglichen Kontexten) vergrößert.[12]

3. Echte und scheinbare Ursachen? In der Literatur, die sich mit probabilistischen Kausalitätsbegriffen beschäftigt, wird es oft für erforderlich gehalten, echte und scheinbare Ursachen zu unterscheiden. Die Idee besteht darin, dass eine Veränderung nur eine scheinbare Ursache ist, wenn sie zur Erklärung eines ihr zugerechneten Effekts bei einer Berücksichtigung weiterer Bedingungen nicht mehr erforderlich ist.[13]

[9] Wie bei stochastischen Kovariablen auch durchschnittliche Effekte definiert werden können, so dass stochastische Varianten einer Kontextabhängigkeit reflektierbar werden, wird in Abschnitt 6.3 besprochen.

[10] Dadurch werden auch automatisch einige der Probleme vermieden werden, die bei Definitionen probabilistischer Kausalität auftreten, wenn unspezifisch auf „Faktoren“ Bezug genommen wird; man vgl. die Diskussion bei Hitchcock (1993).

[11] Solche Versuche findet man etwa bei Eells und Sober (1983: 37), man vgl. auch Eells (1991: 94ff.). Dagegen betont Carroll (1991) anhand von Beispielen die Bedeutung von Kontextabhängigkeiten.

[12] Dies wurde in den ersten Definitionen probabilistischer Kausalität von Suppes (1970) angenommen, an dessen Überlegungen sich die spätere Diskussion lange Zeit orientiert hat, und wird beispielsweise von Sober (1986: 97) gefordert. Weitere Bemerkungen zu dieser Diskussion findet man bei Eells (1988, 1991).

[13] Diese Idee spielte insbesondere bei Suppes (1970) und der sich anschließenden Diskussion über probabilistische Kausalität eine zentrale Rolle. Man findet sie indessen auch in einer Variante des statistischen Kausalitätsbegriffs, die von Cox (1992: 293) folgendermaßen formuliert wird: „[...] a variable x_C is a cause of the response y_E if it occurs in all regression equations for y_E whatever other variables x_B are included.“ Cox diskutiert auch eine Reihe von Problemen, die mit dieser Idee statistischer Kausalität verbunden sind. Eine kritische Diskussion aus soziologischer Perspektive gibt Goldthorpe (2001).

Bei funktionalen Ursachen, wie sie in § 2 definiert worden sind, kann die Unterscheidung jedoch nicht sinnvoll angewendet werden; in dem Kontext, auf den bei der Definition Bezug genommen wird, handelt es sich stets um „echte“ Ursachen. Natürlich muss bedacht werden, dass die Definition ein jeweils bestimmtes Modell voraussetzt; ob eine Veränderung $\Delta(x', x'')$ funktionale Ursache eines stochastischen Effekts $\Delta^s(\dot{Y}; x', x'')$ ist, hängt also auch vom jeweils vorausgesetzten Modell ab. Aber das ist eigentlich eine Selbstverständlichkeit, da sich Aussagen über funktionale Kausalzusammenhänge immer nur auf Modelle beziehen können.[14]

4. Singuläre und generische Kausalaussagen. Man kann singuläre und generische Kausalaussagen unterscheiden.[15] Singuläre Kausalaussagen beziehen sich auf bestimmte Vorkommnisse in der menschlichen Erfahrungswelt; zum Beispiel: Der Aufprall des Steins auf die Fensterscheibe war die Ursache dafür, dass sie zerbrach. Dabei wird ein räumlich und zeitlich bestimmter, empirisch identifizierbarer Kontext vorausgesetzt, in dem von bestimmten Ereignissen gesprochen werden kann. Dagegen beziehen sich generische Kausalaussagen auf Ereignistypen; zum Beispiel: Wenn eine Fensterscheibe von einem Stein getroffen wird, wird sie (sehr wahrscheinlich) zerbrechen. Dies ist eine allgemeine Aussage, bei der nicht auf bestimmte Ereignisse, sondern unspezifisch auf irgendeinen Stein und irgendeine Fensterscheibe Bezug genommen wird. Funktionale Kausalaussagen, wie sie in den ersten beiden Paragraphen dieses Abschnitts definiert worden sind, sind stets generische Kausalaussagen.

Ein wesentlicher Unterschied besteht darin, dass sich singuläre Kausalaussagen unmittelbar auf Aspekte der menschlichen Erfahrungswelt, generische Kausalaussagen auf Modelle beziehen.[16] Dem entspricht eine unterschiedliche Vergegenwärtigung des Kontextes. Bei singulären Kausalaussagen wird ein räumlich und zeitlich bestimmter Kontext vorausgesetzt, in dessen Rahmen sich die Ereignisse abspielen, die durch eine Kausalaussage verknüpft werden. Dieser Kontext lässt sich stets durch beliebig viele Aspekte charakterisieren; solche Aspekte explizit zu benennen, ist jedoch nur erforderlich, um die Kausalaussage verständlich zu machen.

[14] Allerdings findet man in der Literatur auch andere Auffassungen; man vgl. etwa Rosenbaum und Rubin (1984: 27).

[15] In der englischsprachigen Literatur wird auch von *token-level* und *type-level causation* gesprochen (Eells 1991).

[16] Auch wenn man Annahmen über bestimmte Werte von Modellvariablen macht, verbleibt man offenbar in der jeweiligen Modellwelt und erzeugt keine empirischen Tatsachen. Es ist deshalb unklar, was mit folgender Aussage gemeint sein könnte: „I take token causation to be a relation among facts about the values of variables.“ (Hausman 1993: 437)

Anders verhält es sich bei generischen Kausalaussagen. Da bei ihrer Formulierung kein empirisch bestimmter Kontext vorausgesetzt werden kann, muss man sich bemühen, den Kontext, von dem ihre Geltung abhängt, in allgemeiner Form anzugeben, d.h. in Gestalt von Bedingungen, die in einem Modell durch Werte von Variablen angebbar sind. Da stets nur wenige Bedingungen explizit angegeben werden können, bleibt der Geltungsanspruch unvermeidlich auf das Modell beschränkt und kann nicht ohne weiteres auf empirische Kontexte übertragen werden. Es ist deshalb nicht möglich, singuläre aus generischen Kausalaussagen abzuleiten. Zwar können generische Kausalaussagen oft als Regeln verstanden werden, mit deren Hilfe sich singuläre Kausalaussagen formulieren lassen. Dabei muss jedoch der jeweils konkrete Kontext berücksichtigt werden, der in dem Modell, auf das sich die generische Kausalaussage bezieht, bestenfalls partiell erfasst wird.

Ein weiteres Problem entsteht, wenn aus generischen Kausalaussagen, die funktionale Ursachen mit stochastischen Effekten verknüpfen, singuläre Kausalaussagen gewonnen werden sollen. Denn zur Formulierung der singulären Kausalaussagen müssen dann subjektive Wahrscheinlichkeiten für konkrete Ereignisse verwendet werden.

5. Retrospektive und prospektive Fragestellungen. Eine weitere Unterscheidung betrifft die Art der Fragestellung. Retrospektive Fragestellungen beziehen sich darauf, wie (durch welche Ursachen) ein Sachverhalt entstanden ist. Dabei wird bereits vorausgesetzt, dass der zu erklärende Sachverhalt entstanden ist; fraglich ist, ob bzw. wie er als Wirkung bestimmter Ursachen erklärt werden kann. Davon zu unterscheiden sind prospektive Fragestellungen, bei denen nach den wahrscheinlichen Folgen einer bestimmten Ursache gefragt wird. Dabei wird hypothetisch angenommen oder als Tatsache vorausgesetzt, dass eine bestimmte Ursache realisiert worden ist; fraglich ist, welche Wirkungen daraus wahrscheinlich entstehen werden.

Von einigen Autoren ist behauptet worden, dass nur prospektive Kausalfragen eindeutig formuliert und beantwortet werden können;[17] aber ich glaube nicht, dass das richtig ist. Denn wie in Abschnitt 6.3 anhand von Beispielen gezeigt wird, können auch prospektive Kausalfragen nicht immer eindeutig beantwortet werden; und andererseits ist es im Rahmen von Modellen oft möglich, auch retrospektive Kausalfragen mehr oder weniger eindeutig zu beantworten. Weiß man zum Beispiel bei dem in Abschnitt 5.1 (§ 1) betrachteten Modell, dass die Batterie Strom liefert, kommt als Ursache dafür, dass die Glühlampe nicht mehr leuchtet, nur infrage, dass der Schalter geöffnet worden ist.

Richtig ist allerdings, dass retrospektive Kausalfragen oft keine oder nur sehr problematische Antworten erlauben. Denn meistens sind unterschiedliche Ursachen vorstellbar, die einen bestimmten Sachverhalt bewirkt

[17] P. W. Holland (1986: 959; 1988a; 1993).

haben können. Dies gilt insbesondere dann, wenn sich die Fragestellung auf reale Vorkommnisse in der menschlichen Erfahrungswelt bezieht, so dass nicht ohne weiteres ein bestimmtes Modell unterstellt werden kann.

6. Dynamische und komparative Ursachen. Schließlich soll noch einmal betont werden, dass der formale Begriff einer funktionalen Ursache mit unterschiedlichen Vorstellungen über Ursachen verbunden werden kann. Insbesondere umfasst er sowohl dynamische als auch komparative Ursachen. Zur Erläuterung beziehen wir uns auf eine Veränderung $\Delta(x', x'')$ bei einer Modellvariablen $\ddot{X}$.

- Von einer *dynamischen Ursache* kann man sprechen, wenn man sich ein Ereignis vorstellen kann, in dem sich der Wert der Variablen $\ddot{X}$ von x' zu x'' verändert. Zum Beispiel: Ein Schalter wird geschlossen; eine Person wird arbeitslos; ein Patient bekommt ein bestimmtes Medikament.
- Von einer *komparativen Ursache* wird gesprochen, wenn sich die Werte x' und x'' auf zwei unterschiedliche Situationen (oder Objekte) $\sigma(x')$ und $\sigma(x'')$ beziehen, ohne dass man sinnvoll von einem Ereignis sprechen kann, durch das $\sigma(x'')$ aus $\sigma(x')$ hervorgegangen ist.

Sozialwissenschaftliche Untersuchungen beschäftigen sich überwiegend mit komparativen Ursachen. Die Vorgehensweise besteht darin, Menschen oder Situationen, die unterschiedliche Eigenschaften aufweisen, zu vergleichen. So wird zum Beispiel untersucht, wie sich die Arbeitsverdienste von Personen mit unterschiedlichen Ausbildungsniveaus unterscheiden; und dann werden Differenzen im Ausbildungsniveau als eine Ursache für unterschiedliche Arbeitsverdienste betrachtet. Offenbar handelt es sich dann um eine komparative Ursache.

6.2 Ambivalente Bezugnahmen auf Individuen

Unsere Definition funktionaler Kausalität bezieht sich auf funktionale Modelle, die (wie in Abschnitt 5.1 (§ 5) ausgeführt wurde) als Ablaufschemas verstanden werden können. Solche Modelle können zwar unter Umständen verwendet werden, um kausale Vermutungen über bestimmte Individuen bzw. konkrete Situationen zu gewinnen; die Modelle selbst enthalten jedoch nur generische Variablen. Weder erfordern, noch erlauben sie eine Definition kausaler Effekte, die sich auf identifizierbare Individuen bezieht. Hauptsächlich dadurch unterscheidet sich unser Vorschlag zur Definition funktionaler Kausalität von einer anderen Konzeption, die insbesondere durch Beiträge von D. B. Rubin, P. W. Holland und P. R. Rosenbaum angeregt wurde und seither eine weite Verbreitung gefunden hat.[18] In diesem Abschnitt soll gezeigt werden, dass einige der Schwierigkeiten, die mit dem

[18] Man vgl. etwa Holland (1986, 1988, 1994, 2001), Rubin (1974), Rosenbaum und Rubin (1983), Sobel (1995). Zur Verwendung in den Sozialwissenschaften wurde der Ansatz u.a. von King, Keohane und Verba (1994) und Winship und Morgan (1999) propagiert.

Ansatz verbunden sind, vermieden werden können, wenn auf die Annahme identifizierbarer Individuen verzichtet wird. (Weitere Überlegungen zu diesem Ansatz folgen in den Abschnitten 7.2 und ??.)

1. Der Modellansatz von P. W. Holland. Wir beziehen uns auf den Ansatz in einer Variante, die von P. W. Holland entwickelt worden ist. Holland beginnt mit einer Bezugnahme auf eine bestimmte endliche Gesamtheit, in unserer Notation: $\Omega = \{\omega_1, \ldots, \omega_n\}$; die Elemente werden in einer allgemeinen Bedeutung Individuen genannt (es kann sich um beliebige Objekte oder Situationen handeln). Für die weiteren Erläuterungen wird ein experimenteller oder quasi-experimenteller Kontext vorausgesetzt. Um mögliche Ursachen kausaler Effekte zu erfassen, wird ein Merkmalsraum $\tilde{\mathcal{X}} = \{\tilde{x}_1, \ldots, \tilde{x}_m\}$ angenommen. Die Elemente bezeichnen unterschiedliche Behandlungen oder Einflüsse, denen die Individuen in Ω ausgesetzt werden können; eine irgendwie realisierte Zuordnung kann also durch eine statistische Variable

$$X : \Omega \longrightarrow \tilde{\mathcal{X}}$$

erfasst werden. Außerdem wird zur Erfassung möglicher Effekte ein Merkmalsraum $\tilde{\mathcal{Y}}$ angenommen, der hier nicht näher spezifiziert zu werden braucht. Hollands Grundgedanke besteht nun darin, für jeden Wert $\tilde{x} \in \tilde{\mathcal{X}}$ eine statistische Variable

$$Y_{\tilde{x}} : \Omega \longrightarrow \tilde{\mathcal{Y}}$$

anzunehmen, für die er folgende Interpretation vorschlägt: $Y_{\tilde{x}}(\omega)$ ist das Ergebnis, das resultieren würde, wenn das Individuum ω dem Einfluss $\tilde{x}$ ausgesetzt würde. Holland unterstellt also einen deterministischen Zusammenhang zwischen möglichen Ursachen und Wirkungen; aber das ist nicht wesentlich. Von entscheidender Bedeutung ist jedoch, dass Holland annimmt, dass für jedes Individuum $\omega \in \Omega$ alle *möglichen* Wirkungen

$$Y_{\tilde{x}_1}(\omega), \ldots, Y_{\tilde{x}_m}(\omega)$$

in einer irgendwie bestimmten Weise simultan existieren und infolgedessen für sich anschließende theoretische Definitionen verwendet werden können.

Auf den ersten Blick erscheint diese Annahme, insbesondere infolge der von Holland verwendeten „kontrafaktischen“ Sprechweisen, fragwürdig.[19] Die zentrale Annahme, von der Hollands Ansatz ausgeht, kann indessen auch folgendermaßen formuliert werden:

Annahme: Für jedes Individuum $\omega \in \Omega$ existiert eine Funktion: $f_\omega : \tilde{\mathcal{X}} \longrightarrow \tilde{\mathcal{Y}}$ (6.1)

[19]Man vgl. hierzu den Beitrag von Dawid (2000) und die sich anschließende Diskussion sowie Dawid (2006).

Sind solche Funktionen gegeben, liefern sie sogleich die Werte der von Holland supponierten statistischen Variablen: $Y_{\tilde{x}}(\omega) = f_\omega(\tilde{x})$. Zu beachten ist natürlich, dass es sich nicht um eine empirische Annahme handelt, sondern dass durch die Annahme für jedes Individuum aus Ω ein separates Modell unterstellt wird.[20]

Infolge der Annahme (6.1) entsteht die Möglichkeit, *individuenspezifische* kausale Effekte zu definieren:

$$\Delta^i(Y; x', x''; \omega) := f_\omega(x'') - f_\omega(x') = Y_{x''}(\omega) - Y_{x'}(\omega) \tag{6.2}$$

wobei x' und x'' irgendwelche unterschiedlichen Werte in $\tilde{\mathcal{X}}$ sein können. Aus dieser Definition folgt für Holland (1986: 947) das „fundamental problem of causal inference“: dass Werte so definierter Effekte grundsätzlich nicht beobachtet werden können.

2. Konkrete und generische Fragestellungen. Um zu verstehen, worin das Problem eigentlich besteht, ist es sinnvoll, ähnlich wie zwischen singulären und generischen Kausalaussagen zwei Arten kausaler Fragestellungen zu unterscheiden. Erstens konkrete kausale Fragestellungen, die sich darauf beziehen, was in einer konkreten Situation infolge eines bestimmten Ereignisses vermutlich geschehen wird oder wie sich ein identifizierbares Individuum infolge einer bestimmten Veränderung von Bedingungen vermutlich verhalten wird. Es handelt sich um modale Fragestellungen, zum Beispiel: Wie wird sich die Krankheit eines bestimmten Patienten vermutlich entwickeln, wenn ihm ein bestimmtes Medikament gegeben wird?[21]

Offenbar vergleichen Fragen dieser Art zwei oder mehr Möglichkeiten, und man kann bestenfalls beobachten, was bei einer – der jeweils realisierten Möglichkeit – herausgekommen ist. Klar ist aber auch, dass es sich gar nicht um ein Beobachtungsproblem handelt; das Problem besteht vielmehr darin, den unterschiedlichen Möglichkeiten entsprechende konditionale Aussagen zu begründen. Das dafür erforderliche Wissen kann zwar nur aus Beobachtungen gewonnen werden, diese Beobachtungen können sich aber gar nicht auf diejenige Situation beziehen, auf die sich die konkrete kausale Fragestellung bezieht. Sie können vielmehr nur aus vergleichbaren Situationen stammen.

Damit beginnt die Aufgabe, Modelle zu entwickeln, durch die Beobachtungen aus vergleichbaren Situationen nutzbar gemacht werden können.

[20]Diese funktionalen Modelle werden von Holland (2001: 178) auch als „kausale Modelle“ bezeichnet. Im Ansatz von Holland handelt es sich um deterministische Modelle; ohne wesentliche Änderungen könnten aber auch stochastische Modelle verwendet werden. Zum Beispiel unterstellt die von King, Keohane und Verba (1994) verwendete Rhetorik stochastische Modelle.

[21]Modale Fragestellungen beziehen sich allgemein auf Möglichkeiten und setzen nicht unbedingt kontrafaktische Unterstellungen voraus. Dies kann aber der Fall sein, wie zum Beispiel bei der Fragestellung: Wie hätte sich die Krankheit des Patienten entwickelt, wenn ihm kein (oder ein anderes) Medikament gegeben worden wäre? Man vgl. zu diesen Varianten modaler Fragestellungen auch Dawid (2000; 2006).

Damit verändert sich indessen auch die Art der Fragestellung. Die Modellbildung muss von generischen kausalen Fragestellungen ausgehen, die sich auf Situationen bzw. Individuen beziehen, die in allgemeiner Weise durch Werte von Variablen charakterisiert werden können. Dann wird aber auch Hollands Problem in gewisser Weise gegenstandslos, wie ein Vergleich der folgenden Modellansätze zeigt:

$$\text{(a)}\quad \begin{tikzcd}[row sep=small] \ddot{X} \arrow[dr] & \\ & \ddot{Y} \\ \ddot{I} \arrow[ur] & \end{tikzcd} \qquad\qquad \text{(b)}\quad \begin{tikzcd}[row sep=small] \ddot{X} \arrow[dr] & \\ & \dot{Y} \\ \ddot{Z} \arrow[ur] & \end{tikzcd} \qquad (6.3)$$

Hollands Ansatz entspricht der Variante (a). $\ddot{X}$ erfasst Werte aus dem Wertebereich $\tilde{\mathcal{X}}$ (Einflüsse, denen ein Individuum ausgesetzt werden kann), $\ddot{Y}$ mit dem Wertebereich $\tilde{\mathcal{Y}}$ erfasst die Reaktionen. Außerdem gibt es eine Modellvariable $\ddot{I}$, die dazu dienen soll, auf Individuen der vorausgesetzten Gesamtheit Ω Bezug zu nehmen. Es ist eine Modellvariable mit dem Wertebereich Ω.[22]

Offenbar geht diese Variante nicht von einer generischen Fragestellung aus, denn sie bezieht sich von vornherein nur auf die Individuen aus der Menge Ω. Zugleich eignet sie sich aber auch nicht für konkrete kausale Fragestellungen, die sich auf ein bestimmtes Individuum ω^* beziehen. Denn da Holland annimmt, dass für jedes Individuum in Ω ein bestimmter Wert von $\ddot{X}$ und $\ddot{Y}$ realisiert worden ist, kann ω^* kein Element von Ω sein. Dann ist jedoch unklar, wie mithilfe des Modells konkrete kausale Fragestellungen für ω^* beantwortet werden könnten.[23]

Dagegen orientiert sich die Modellvariante (b) an einer generischen Fragestellung. Zu diesem Zweck wird eine statistische Betrachtungsweise eingenommen, d.h. es wird davon ausgegangen, dass man Individuen nur durch Werte von Variablen unterscheiden kann, wobei die Wertebereiche der Variablen ohne Verwendung von Namen für Individuen definierbar sein müssen.[24] Für die Modellbildung kann eine beliebige (auch mehrdimensionale) Variable $\ddot{Z}$ mit einem Wertebereich $\tilde{\mathcal{Z}}$ angenommen werden. Das Modell bezieht sich dann nicht auf Individuen einer bestimmten Menge Ω, sondern in unbestimmter Weise auf Individuen, die durch Werte von $\ddot{Z}$ charakterisiert werden können. Aus der Sicht des Modells werden also Individuen mit identischen Werten der Variablen $\ddot{Z}$ nicht unterschieden. Infolgedessen muss damit gerechnet werden, dass trotz gleicher Werte von

[22] Eine Variante dieses Modells, bei der anstelle von $\ddot{X}$ und $\ddot{I}$ Zufallsvariablen verwendet werden, findet man bei Steyer et al. (2000).

[23] Eine Überlegung besteht darin, dass Ω eine „repräsentative Stichprobe" aus einer Superpopulation Ω^* und ω^* ein Element aus Ω^* ist. Diese Überlegung betrifft aber bestenfalls die Anwendungsbedingungen des Modells, nicht die Bedeutung der durch das Modell definierbaren Größen. Wir werden die Überlegung später genauer diskutieren.

[24] Insbesondere sind dann Funktionen, die jedem Element einer statistischen Gesamtheit eine Identifikationsnummer zuweisen, keine zulässigen Variablen.

$\ddot{X}$ und $\ddot{Z}$ unterschiedliche Werte im Merkmalsraum $\tilde{\mathcal{Y}}$ realisiert werden können. In der Modellformulierung wird dem dadurch Rechnung getragen, dass die abhängige Variable als eine stochastische Variable $\dot{Y}$ konzipiert und durch eine stochastische Funktion von $\ddot{X}$ und $\ddot{Z}$ abhängig gemacht wird.

3. Unterschiedliche Homogenitätsannahmen. Eine wichtige Rolle in der Diskussion über Bedingungen, unter denen kausale Effekte mithilfe von Daten geschätzt werden können, spielt die Annahme individueller Homogenität. Im Ansatz von Holland besteht sie in der Annahme, dass die Funktionen f_ω für alle Individuen in Ω (oder in Subgruppen, die durch generische Variablen definiert sind) identisch sind.[25] Etwas andere Formulierungen sind erforderlich, wenn man für die Individuen stochastische Funktionen unterstellt; zum Beispiel:

> „For a data set with N observations, unit homogeneity is the assumption that all units with the same value of the explanatory variables have the same expected value of the dependent variable." (King et al. 1994: 91)

Ob eine deterministische oder eine stochastische Formulierung verwendet wird, ist jedoch nicht entscheidend. Entscheidend ist vielmehr, dass in beiden Varianten auf identifizierbare Individuen Bezug genommen wird. Die Annahme besteht darin, dass sich alle Individuen einer Gruppe unter gleichen Bedingungen auf gleiche Weise verhalten.

Eine solche Annahme ist indessen nicht nur grundsätzlich fragwürdig; sie widerspricht auch einer statistischen Betrachtungsweise. Denn eine statistische Betrachtungsweise setzt nicht voraus, dass man annimmt, dass alle Untersuchungseinheiten im Rahmen der jeweils verwendeten Klassifikationen gleich sind oder sich gleich verhalten; sondern sie geht von der Idee aus, dass es (als Bedingung für die Anwendbarkeit einer statistischen Betrachtungsweise) unter Umständen sinnvoll sein kann, nur bestimmte Unterschiede zwischen den Untersuchungseinheiten zu berücksichtigen und von allen übrigen Unterschieden, die es sicherlich gibt, zu abstrahieren.

Dieser Betrachtungsweise entspricht das Modell (6.3b). Offenbar ist es nicht erforderlich anzunehmen, dass die Funktionen f_ω bei allen Individuen, die den gleichen Wert von $\ddot{Z}$ aufweisen, identisch sind. Tatsächlich setzt das Modell nicht einmal die Annahme voraus, dass man über solche Funktionen sinnvoll sprechen kann (also die Annahme (6.1)). Stattdessen impliziert das Modell in gewisser Weise eine statistische Homogenitätsannahme (obwohl man besser von einer Abstraktion sprechen sollte): dass es für jeden Wert $\ddot{Z} = z$ eine bestimmte stochastische Funktion

$$x \longrightarrow \Pr[\dot{Y} \mid \ddot{X} = x, \ddot{Z} = z]$$

gibt. Hier kann dann die Idee ansetzen, Werte von $\ddot{Z}$ als Kovariablenkontexte für funktionale Kausalbeziehungen zwischen $\ddot{X}$ und $\dot{Y}$ zu betrachten.

[25] Vgl. Holland (1986: 948).

4. Generische Kausaleffekte für Individuen. Es ist bemerkenswert, dass Variablen der Form $Y_{\tilde{x}}$ (vgl. § 1) ausgehend vom Modell (6.3b) nicht definiert werden können. Stattdessen könnten Variablen der Form

$$Y^*_{\tilde{x}} : \Omega \longrightarrow \mathbf{R}$$

betrachtet werden, deren Werte durch $Y^*_{\tilde{x}}(\omega) := \mathrm{E}(\dot{Y} \mid \ddot{X} = \tilde{x}, \ddot{Z} = Z(\omega))$ definiert sind.[26] Somit könnten auch generische Kausaleffekte für Individuen definiert werden:

$$\Delta^*(\dot{Y}; x', x''; \omega) := Y^*_{x''}(\omega) - Y^*_{x'}(\omega)$$

Im Unterschied zu den in (6.2) definierten Effekten handelt es sich jetzt jedoch um stochastische Effekte, die jeweils für Klassen von Individuen (die durch $\ddot{Z}$ induzierten Äquivalenzklassen) definiert sind. Zwar könnten diese stochastischen Effekte ausgehend von einer endlichen Gesamtheit Ω und der Annahme (6.1) auch als Durchschnittswerte der in (6.2) definierten deterministischen Effekte aufgefasst werden:

$$\Delta^*(\dot{Y}; x', x''; \omega) = \frac{1}{|\{\omega' | \ddot{Z}(\omega') = \ddot{Z}(\omega)\}|} \sum_{\omega' \in \{\omega' | \ddot{Z}(\omega') = \ddot{Z}(\omega)\}} \Delta^i(Y; x', x''; \omega')$$

Diese durch die Annahme (6.1) ermöglichte Interpretation ist jedoch nicht erforderlich, um den stochastischen Effekten eine kausale Bedeutung zu geben. Und insofern sie eine bestimmte Gesamtheit von Individuen (Ω) voraussetzt, ist sie auch weder mit konkreten noch mit generischen kausalen Fragestellungen vereinbar.

6.3 Isolierbarkeit funktionaler Ursachen

Mit einem Reden von (funktionalen) Ursachen verbindet sich meistens die Vorstellung, dass ihnen bestimmte Wirkungen zurechenbar sind.[27] Die in Abschnitt 6.1 eingeführten Definitionen entsprechen dieser Vorstellung, indem sie das Reden von funktionalen Ursachen bzw. Wirkungen, soweit erforderlich, von Kovariablenkontexten abhängig machen. Die Definitionen setzen deshalb voraus, dass bei einer Veränderung $\Delta(x', x'')$ des Werts einer Variablen ein ggf. relevanter Kovariablenkontext konstantgehalten werden kann.[28] Das ist jedoch nicht immer möglich, insbesondere bei endogenen Variablen eines Modells. In diesem Abschnitt werden einige der

[26] Als Wertebereich werden die reellen Zahlen $\mathbf{R}$ verwendet, da Erwartungswerte nicht unbedingt Elemente des Wertebereichs $\tilde{\mathcal{Y}}$ von $\dot{Y}$ sein müssen.

[27] Man vgl. beispielsweise McKim (1997: 7f.).

[28] Darauf weist auch Heckman (2000: 52f.) in seiner Definition funktionaler Kausalität ausdrücklich hin; ebenso Heckmann (2005: 1): „Holding all factors save one at a constant level, the change in the outcome associated with manipulation of the varied factor is called a causal effect of the manipulated factor.“

daraus resultierenden Schwierigkeiten besprochen.

1. Deterministische Kovariablen. Zunächst soll noch einmal rekapituliert werden, wie Kontextabhängigkeit bei deterministischen Kovariablen berücksichtigt werden kann. Einem Modell, in dem die Verteilung einer stochastischen Variablen $\dot{Y}$ von Werten deterministischer Variablen $\ddot{X}$ und $\ddot{Z}$ abhängt, entspricht eine stochastische Funktion

$$(x, z) \longrightarrow \Pr[\dot{Y} | \ddot{X} = x, \ddot{Z} = z] \tag{6.4}$$

Stochastische Effekte einer Veränderung $\Delta(x', x'')$ bei der Variablen $\ddot{X}$ können dann beispielsweise durch

$$\Delta^s(\dot{Y}; x', x''; z) := \mathrm{E}(\dot{Y} | \ddot{X} = x'', \ddot{Z} = z) - \mathrm{E}(\dot{Y} | \ddot{X} = x', \ddot{Z} = z) \tag{6.5}$$

definiert werden, so dass unmittelbar sichtbar wird, ob bzw. wie der stochastische Effekt auch vom Kovariablenkontext $\ddot{Z} = z$ abhängig ist.

Ist der Effekt kontextabhängig, könnte man versuchen, einen durchschnittlichen Effekt zu berechnen, indem man Annahmen über eine Verteilung der Werte von $\ddot{Z}$ (in irgendeiner Population) macht. Aber solche Annahmen stehen offenbar innerhalb des Modells nicht zur Verfügung und setzen ein erweitertes Populationsmodell voraus.[29]

2. Abhängige Kovariablen. Bei dem Modell in § 1 wurde angenommen, dass bei einer Veränderung $\Delta(x', x'')$ des Werts von $\ddot{X}$ ein durch $\ddot{Z} = z$ bestimmter Kovariablenkontext konstantgehalten werden kann. Man kann sich auch Modelle vorstellen, bei denen diese Bedingung, die erforderlich ist, um $\Delta(x', x'')$ als eine isolierbare funktionale Ursache auffassen zu können, nicht oder nur eingeschränkt gegeben ist. Einschränkungen können zum Beispiel aus Constraints resultieren, die die Möglichkeiten, $\ddot{X}$ und $\ddot{Z}$ unabhängig voneinander zu variieren, einschränken. Eine weitere Möglichkeit besteht darin, dass $\ddot{Z}$ eine endogene Modellvariable ist, die durch eine deterministische Funktion von $\ddot{X}$ abhängig ist. Das Modell sieht dann folgendermaßen aus:

$$\begin{tikzcd} \ddot{X} \arrow[d] \arrow[dr] & \\ \ddot{Z} \arrow[r] & \dot{Y} \end{tikzcd} \tag{6.6}$$

Außer der stochastischen Funktion (6.4) gibt es dann noch eine deterministische Funktion $z = g(x)$, so dass die Verteilung von $\dot{Y}$ nur von Werten von $\ddot{X}$ abhängig ist:

$$x \longrightarrow \Pr[\dot{Y} | \ddot{X} = x, \ddot{Z} = g(x)]$$

[29] Das wird ausführlich in Kapitel **??** besprochen.

Zwar kann bei einer Veränderung $\Delta(x', x'')$ des Werts von $\ddot{X}$ der Wert von $\ddot{Z}$ nicht konstantgehalten werden; gleichwohl kann $\Delta(x', x'')$ als eine funktionale Ursache eines stochastischen Effekts

$$\mathrm{E}(\dot{Y}|\ddot{X} = x'', \ddot{Z} = g(x'')) - \mathrm{E}(\dot{Y}|\ddot{X} = x', \ddot{Z} = g(x'))$$

aufgefasst werden. Denn da $\ddot{Z}$ auf deterministische Weise von $\ddot{X}$ abhängt, brauchen Werte von $\ddot{Z}$ nicht als Kovariablenkontexte für Effekte von Veränderungen der Werte von $\ddot{X}$ betrachtet zu werden. Aber das bedeutet natürlich auch, dass Veränderungen der Werte von $\ddot{Z}$ in diesem Modell nicht als isolierbare funktionale Ursachen eines für $\dot{Y}$ definierten stochastischen Effekts betrachtet werden können.

3. Stochastische Kovariablen. Jetzt nehmen wir an, dass der Kovariablenkontext nicht durch eine deterministische Variable $\ddot{Z}$, sondern durch eine stochastische Variable $\dot{Z}$ gegeben ist. Zunächst soll auch angenommen werden, dass sowohl $\ddot{X}$ als auch $\dot{Z}$ exogene Variablen sind, so dass das Modell folgendermaßen aussieht:

$$\begin{tikzcd}[row sep=small] \ddot{X} \arrow[dr] & \\ & \dot{Y} \\ \dot{Z} \arrow[ur] & \end{tikzcd} \qquad (6.7)$$

Es wird angenommen, dass $\dot{Z}$ von $\ddot{X}$ stochastisch unabhängig ist und eine bestimmte Verteilung $\Pr[\dot{Z}]$ hat. Man kann also durch

$$\Delta^s(\dot{Y}; x', x''; \dot{Z}) := \sum_z \left(\mathrm{E}(\dot{Y}|\ddot{X} = x'', \dot{Z} = z) - \mathrm{E}(\dot{Y}|\ddot{X} = x', \dot{Z} = z)\right) \Pr(\dot{Z} = z) \qquad (6.8)$$

einen bzgl. der Verteilung von $\dot{Z}$ durchschnittlichen stochastischen Effekt einer Veränderung $\Delta(x', x'')$ des Werts von $\ddot{X}$ definieren. Er entspricht dem Effekt von $\Delta(x', x'')$ in dem aus (6.7) ableitbaren reduzierten Modell.

4. Interaktion und Verteilungsabhängigkeit. Offenbar kann der in (6.8) definierte stochastische Effekt auch von der Verteilung von $\dot{Z}$ abhängig sein. Dies ist dann der Fall, wenn $\ddot{X}$ und $\dot{Z}$ interaktive Bedingungen sind. Wir verwenden folgende Definitionen für deterministische und stochastische Kovariablen:

- $\ddot{X}$ und $\ddot{Z}$ sind *interaktive Bedingungen* für $\dot{Y}$, wenn der durch (6.5) definierte stochastische Effekt $\Delta^s(\dot{Y}; x', x''; z)$ von z abhängig ist; wenn es also mindestens zwei Werte z' und z'' gibt, so dass $\Delta^s(\dot{Y}; x', x''; z') \neq \Delta^s(\dot{Y}; x', x''; z'')$ ist.
- Wenn $\dot{Z}$ eine stochastische Variable ist, kann eine entsprechende Definition verwendet werden, indem in der Definition (6.5) $\ddot{Z}$ durch $\dot{Z}$ ersetzt wird. Außerdem gibt es folgende Möglichkeit: Der durch (6.8) definierte stochastische Effekt ist *verteilungsabhängig*, wenn er von der Verteilung $\Pr[\dot{Z}]$ abhängt.[30]

Die Definitionen können anhand linearer Regressionsfunktionen verdeutlicht werden. Bei einer Regressionsfunktion

$$\mathrm{E}(\dot{Y}|\ddot{X} = x, \dot{Z} = z) = \alpha + x\beta_x + z\beta_z$$

sind $\ddot{X}$ und $\dot{Z}$ keine interaktiven Bedingungen und der Effekt von $\ddot{X}$ ist auch nicht von der Verteilung von $\dot{Z}$ abhängig. Wird dann durch

$$\mathrm{E}(\dot{Y}|\ddot{X} = x, \dot{Z} = z) = \alpha + x\beta_x + z\beta_z + xz\beta_{xz}$$

ein Interaktionseffekt eingeführt, werden $\ddot{X}$ und $\dot{Z}$ zu interaktiven Bedingungen und der stochastische Effekt wird verteilungsabhängig:

$$\Delta^s(\dot{Y}; x', x''; \dot{Z}) = (x'' - x')(\beta_x + \beta_{xz}\,\mathrm{E}(\dot{Z}))$$

Folgendes Beispiel liefert eine einfache Illustration: $\ddot{X}$ erfasst das Bildungsniveau der Eltern (0 niedrig, 1 hoch), $\dot{Z}$ erfasst den Schultyp (0 Grundschule, 1 weiterführende Schule), und $\dot{Y}$ erfasst, wie die Schule beendet wurde (0 ohne, 1 mit Erfolg). Es werden folgende bedingten Erwartungswerte angenommen:

x	z	$\mathrm{E}(\dot{Y}\|\ddot{X} = x, \dot{Z} = z)$
0	0	0.8
0	1	0.6
1	0	0.8
1	1	0.9

(6.9)

Für den Erfolg in der Grundschule spielt das Bildungsniveau der Eltern keine Rolle, in der weiterführenden Schule führt jedoch ein hohes Bildungsniveau der Eltern zu einer größeren Wahrscheinlichkeit für einen erfolgreichen Abschluss. Offenbar sind $\ddot{X}$ und $\dot{Z}$ interaktive Bedingungen für den Schulerfolg: $\Delta^s(\dot{Y}; 0, 1; 0) = 0$ und $\Delta^s(\dot{Y}; 0, 1; 1) = 0.3$. Infolgedessen ist der Effekt verteilungsabhängig:

$$\Delta^s(\dot{Y}; 0, 1; \dot{Z}) = \Delta^s(\dot{Y}; 0, 1; 1)\Pr(\dot{Z} = 1)$$

so dass der Zusammenhang zwischen dem Bildungsniveau der Eltern und dem Schulerfolg auch von der Wahrscheinlichkeit für den Besuch einer weiterführenden Schule abhängt.

Gleichwohl kann angenommen werden, dass die Verteilung von $\dot{Z}$ nicht von $\ddot{X}$ abhängt, so dass man beliebige Annahmen über diese Verteilung machen kann, ohne den Modellansatz zu verändern. Wie in § 3 ausgeführt

[30] Damit (6.8) definiert ist, wird gleichwohl vorausgesetzt, dass die Verteilung von $\dot{Z}$ nicht von $\ddot{X}$ abhängt.

wurde, ist das eine Bedingung dafür, um einer Veränderung $\Delta(x', x'')$ einen isolierbaren Effekt zurechnen zu können. Wenn der Effekt verteilungsabhängig ist, muss das indessen in der Formulierung berücksichtigt und explizit gemacht werden. Es handelt sich um eine Variante der Kontextabhängigkeit funktionaler Ursachen bzw. Wirkungen, die bei deterministischen Kovariablen als Abhängigkeit von deren Werten und bei stochastischen Kovariablen als Abhängigkeit von deren Verteilungen dargestellt werden muss.[31]

5. Exogene und endogene Ursachen. Will man eine Veränderung $\Delta(x', x'')$ des Werts einer Variablen als eine mögliche funktionale Ursache betrachten, kann es sich um eine exogene oder um eine endogene Variable handeln. In beiden Fällen kommt die Veränderung $\Delta(x', x'')$ zunächst durch eine Annahme zustande: Es wird angenommen, dass sich der Wert einer Variablen von x' zu x'' verändert, und das setzt zunächst nur voraus, dass beide Werte innerhalb des vorausgesetzten Modells möglich sind. Wenn es sich um eine endogene Variable handelt, muss jedoch bedacht werden, dass ihre Werte oder Verteilungen auch von Werten bzw. Verteilungen anderer Modellvariablen abhängig sind.

Zu unterscheiden ist, ob die endogene Variable deterministisch oder stochastisch ist. Handelt es sich um eine deterministische Variable, kann man beispielsweise folgendes Modell betrachten:

$$\ddot{X} \longrightarrow \ddot{Z} \twoheadrightarrow \dot{Y} \tag{6.10}$$

in dem $\ddot{X}$ und $\ddot{Z}$ durch eine deterministische Funktion $z = g(x)$ verknüpft sind. In diesem Fall erscheint es auf unproblematische Weise möglich, einer Veränderung $\Delta(z', z'')$ des Werts der endogenen Variablen $\ddot{Z}$ einen bestimmten für $\dot{Y}$ definierten stochastischen Effekt zuzurechnen. Denn wenn z' und z'' innerhalb des Modells möglich sind, gibt es auch mindestens zwei Werte x' und x'', so dass $z' = g(x')$ und $z'' = g(x'')$ ist. Anders verhält es sich jedoch in dem in § 2 diskutierten Beispiel, bei dem es nicht möglich erscheint, einer endogenen deterministischen Variablen $\ddot{Z}$ einen bestimmten Effekt zuzurechnen.[32]

Jetzt betrachten wir in Analogie zu (6.10) ein einfaches Modell mit einer stochastischen endogenen Variablen:

$$\ddot{X} \twoheadrightarrow \dot{Z} \twoheadrightarrow \dot{Y} \tag{6.11}$$

Wie im Modell (6.10) kann man einer Veränderung $\Delta(z', z'')$ bei der Variablen $\dot{Z}$ einen bestimmten stochastischen Effekt zurechnen. Es gibt aber

[31] Eine ausführliche Beschäftigung mit Problemen der Verteilungsabhängigkeit erfolgt in Kapitel **??**.

[32] Man kann es auch so sagen: Sowohl im Modell (6.6) als auch im Modell (6.10) ist es nicht möglich, bei Veränderungen $\Delta(z', z'')$ Werte von $\ddot{X}$ konstantzuhalten. Im Modell (6.10) ist dies jedoch nicht erforderlich, da $\ddot{X}$ keinen Kovariablenkontext für $\ddot{Z}$ bildet.

auch einen bemerkenswerten Unterschied: Während man im Modell (6.10) bestimmte Veränderungen $\Delta(z', z'')$ durch Interventionen bei der exogenen Variablen bewirken kann, ist dies im Modell (6.11) nicht möglich. Oder anders formuliert: stochastische endogene Ursachen können nicht bewirkt, sondern nur beobachtet werden.

6. Direkte und indirekte Effekte. Weitere Überlegungen sind erforderlich, wenn $\dot{Y}$ auch noch auf direkte Weise von $\ddot{X}$ abhängt, wenn also ein Modell der folgenden Form betrachtet wird:

$$\begin{tikzcd} \ddot{X} \arrow[d] \arrow[dr, twoheadrightarrow] & \\ \dot{Z} \arrow[r, twoheadrightarrow] & \dot{Y} \end{tikzcd} \tag{6.12}$$

Zur Illustration betrachten wir eine Variante des Beispiels aus § 4, nehmen also an, dass $\dot{Z}$ den Schultyp, $\dot{Y}$ den schulischen Abschluss und $\ddot{X}$ das Bildungsniveau der Eltern erfasst (jeweils 0-1-Variablen). Für die Funktionen werden folgende Annahmen getroffen:

x	z	$\mathrm{E}(\dot{Y}\vert\ddot{X}=x, \dot{Z}=z)$
0	0	0.5
0	1	0.7
1	0	0.8
1	1	0.9

x	$\mathrm{E}(\dot{Z}\vert\ddot{X}=x)$
0	0.4
1	0.8

(6.13)

Im Vergleich zum Modell (6.11) ergeben sich zwei Unterschiede. Erstens sind Effekte von Veränderungen bei der endogenen Variablen $\dot{Z}$ jetzt kontextabhängig, zum Beispiel findet man bei einer Veränderung $\Delta(0, 1)$:

$$\mathrm{E}(\dot{Y}|\dot{Z}=1, \ddot{X}=0) - \mathrm{E}(\dot{Y}|\dot{Z}=0, \ddot{X}=0) = 0.7 - 0.5 = 0.2$$
$$\mathrm{E}(\dot{Y}|\dot{Z}=1, \ddot{X}=1) - \mathrm{E}(\dot{Y}|\dot{Z}=0, \ddot{X}=1) = 0.9 - 0.8 = 0.1$$

Zweitens kann man direkte und indirekte Effekte von Veränderungen $\Delta(x', x'')$ bei der Variablen $\ddot{X}$ unterscheiden. Zunächst kann man einen Gesamteffekt berechnen, bei einer Veränderung $\Delta(0, 1)$:[33]

$$\mathrm{E}(\dot{Y}|\ddot{X}=1) - \mathrm{E}(\dot{Y}|\ddot{X}=0) = 0.88 - 0.58 = 0.3$$

Weiterhin können direkte Effekte berechnet werden, indem man $\dot{Z}$ als einen

[33] Zur Berechnung kann die Gleichung

$$\mathrm{E}(\dot{Y}|\ddot{X}=x) = \\ \mathrm{E}(\dot{Y}|\ddot{X}=x, \dot{Z}=0)\,\Pr(\dot{Z}=0|\ddot{X}=x) + \mathrm{E}(\dot{Y}|\ddot{X}=x, \dot{Z}=1)\,\Pr(\dot{Z}=1|\ddot{X}=x)$$

verwendet werden.

Kovariablenkontext betrachtet und annimmt, dass bestimmte Werte fixiert werden können. In unserem Beispiel erhält man folgende Effekte:

$$\mathrm{E}(\dot{Y}|\ddot{X}=1,\dot{Z}=0)-\mathrm{E}(\dot{Y}|\ddot{X}=0,\dot{Z}=0)=0.8-0.5=0.3$$
$$\mathrm{E}(\dot{Y}|\ddot{X}=1,\dot{Z}=1)-\mathrm{E}(\dot{Y}|\ddot{X}=0,\dot{Z}=1)=0.9-0.7=0.2$$

Offenbar sind sie kontextabhängig, so dass es nicht möglich ist, einen bestimmten (kontextunabhängigen) direkten Effekt (des Bildungsniveaus der Eltern für den Schulerfolg) zu berechnen. Infolgedessen kann aber auch kein bestimmer indirekter Effekt, der aus den unterschiedlichen Wahrscheinlichkeiten für die Wahl eines Schultyps resultiert, angegeben werden. Das Beispiel zeigt also auch, dass eine eindeutige Zerlegung in direkte und indirekte Effekte nicht immer möglich ist.[34]

[34]Einige weitere Überlegungen erfolgen in Abschnitt 7.2 (§ 6).

Kapitel 7

Modelle und statistische Daten

Funktionale Modelle sind theoretische Konstruktionen ohne einen unmittelbaren, durch die Modellformulierung bereits artikulierten Realitätsbezug. Natürlich müssen Realitätsbezüge hergestellt werden, wenn man funktionale Modelle für (retrospektive) Erklärungen oder (prospektive) Prognosen verwenden möchte; und wenn man an quantitativen Aussagen interessiert ist, ist es außerdem erforderlich, die in einer Modellformulierung verwendeten Funktionen numerisch zu spezifizieren. Somit kann es auch sinnvoll sein, die Funktionen eines Modells mithilfe von statistischen Daten zu schätzen.[1]

[1]Wenn im Folgenden gelegentlich die Formulierung 'ein Modell schätzen' verwendet wird, ist gemeint: dass statistische Daten verwendet werden, um Einsichten in die Beschaffenheit der durch das Modell angenommenen Funktionen zu gewinnen. Die Formulierung setzt nicht voraus, aber schließt es auch nicht aus, dass die Modellfunktionen vorab in irgendeine bestimmte parametrische Form gebracht werden.

In diesem Kapitel werden einige der damit verbundenen Probleme besprochen. Dabei geht es nur um konzeptionelle Fragen, nicht darum, wie man parametrische Modelle mit statistischen Daten schätzen kann. Es gibt drei Abschnitte. Im ersten Abschnitt wird gezeigt, dass es keinen unmittelbaren Zusammenhang zwischen funktionalen Modellen und statistischen Daten gibt. Im zweiten Abschnitt werden einige Aspekte der Frage besprochen, wie man mithilfe statistischer Daten Einsichten in funktionale Kausalzusammenhänge gewinnen kann. Dabei wird zunächst, wie in der Literatur üblich, auf randomisierte Experimente Bezug genommen, dann aber gezeigt, dass dieser Ansatz bei sozialwissenschaftlichen Anwendungen auf Sinngrenzen stößt. Schließlich wird im dritten Abschnitt besprochen, dass es unterschiedliche Bezugsprobleme gibt, die auch für empirische Verwendungen funktionaler Modelle relevant sind.

7.1 Funktionale Modelle und Daten

In diesem Abschnitt soll deutlich gemacht werden, dass es keinen unmittelbaren und eindeutigen Zusammenhang zwischen funktionalen Modellen und statistischen Daten gibt.

1. Funktionale Modelle und Datenmodelle. Funktionale Modelle formulieren Zusammenhänge zwischen Modellvariablen und dienen einer Reflexion modaler Fragestellungen. *Datenmodelle* beziehen sich auf Daten: in der Vergangenheit realisierte Sachverhalte, soweit sie durch statistische Variablen erfasst werden können. Wir unterscheiden zwei Varianten; in beiden Fällen wird auf eine (meistens mehrdimensionale) statistische Variable S Bezug genommen, die für eine Gesamtheit Ω definiert ist.

– Ein *statistisches Datenmodell* dient dem Zweck, die statistische Verteilung $\mathrm{P}[S]$ oder bestimmte Aspekte dieser Verteilung (beispielsweise Regressionsfunktionen) darzustellen.

– Ein *stochastisches Datenmodell* formuliert einen Zufallsgenerator, so dass die gegebenen Daten als eine Realisierung betrachtet werden können.[2] Diese Betrachtungsweise erlaubt es dann, den Zufallsgenerator als ein Modell für einen datenerzeugenden (ggf. auch substantiellen) Prozess zu bezeichnen, durch den die Daten (und ggf. die Sachverhalte, auf die sie sich beziehen) entstanden sein könnten.[3]

[2] Es sei angemerkt, dass auch solche Modelle in der Literatur oft als *statistische* Modelle bezeichnet werden. Es gibt keinen einheitlichen Sprachgebrauch. Unsere Terminologie folgt der Unterscheidung zwischen statistischen und stochastischen Variablen.

[3] Von einer solchen Betrachtungsweise wird auch in der statistischen Literatur, soweit sie sich an stochastischen Modellen orientiert, oft ausgegangen. Bei Cox und Wermuth (1996: 12) findet sich folgende Erläuterung: „The basic assumptions of *probabilistic analyses* are as follows: 1. The data are observed values of random variables, i.e. of variables having a probability distribution. 2. Reasonable working assumptions can be made about the nature of these distributions, usually that they are of a particular mathe-

Stochastische Datenmodelle können auch als Hilfsmittel angesehen werden, um funktionale Modelle mit statistischen Daten zu schätzen. Ausgehend von einem funktionalen Modell wird ein stochastisches Datenmodell für die jeweils gegebenen Daten gebildet. Im einfachsten Fall (der für die Illustrationen in diesem Kapitel stets vorausgesetzt wird) können dann die aus den Daten berechenbaren (bedingten) Häufigkeitsverteilungen unmittelbar als Schätzungen der entsprechenden durch das Modell definierten (bedingten) Wahrscheinlichkeitsverteilungen verwendet werden.

2. Datenerzeugung mit funktionalen Modellen. Jetzt soll überlegt werden, wie ausgehend von einem funktionalen Modell $\mathcal{M} = (\mathcal{V}, \mathcal{F})$ ihm entsprechende statistische Daten erzeugt werden können. Es gibt vier Schritte:

(1) Zunächst wird eine Referenzmenge Ω definiert, deren Elemente als Bezugseinheiten für jeweils spezifische Belegungen der Modellvariablen mit Werten dienen sollen.

(2) Dann werden korrespondierend zu den Modellvariablen in $\mathcal{V}$ statistische Variablen für Ω definiert. Enthält $\mathcal{V}$ beispielsweise die Modellvariablen $\ddot{X}$, $\ddot{Z}$ und $\dot{Y}$ mit den Wertebereichen $\tilde{\mathcal{X}}$, $\tilde{\mathcal{Z}}$ und $\tilde{\mathcal{Y}}$, wird eine dreidimensionale statistische Variable

$$(X, Z, Y) : \Omega \longrightarrow \tilde{\mathcal{X}} \times \tilde{\mathcal{Z}} \times \tilde{\mathcal{Y}}$$

definiert und angenommen, dass $(X, Z, Y)(\omega)$ die bei der Bezugseinheit $\omega \in \Omega$ realisierten Werte der Modellvariablen liefert.

(3) Für jede Bezugseinheit $\omega \in \Omega$ werden dann zunächst für diejenigen statistischen Variablen, die den exogenen Modellvariablen entsprechen, bestimmte Werte erzeugt, und zwar: bei stochastischen exogenen Variablen mithilfe von Zufallsgeneratoren und bei deterministischen exogenen Variablen *auf beliebige Weise* (natürlich unter Beachtung ggf. definierter Constraints).

(4) Dann werden ausgehend von den jeweils angenommenen Werten der exogenen Modellvariablen unter Verwendung der durch das Modell vorausgesetzten Funktionen Werte der endogenen Modellvariablen erzeugt und als Werte der entsprechenden statistischen Variablen verwendet. Wenn es sich um stochastische Funktionen handelt, werden zur Erzeugung bestimmter Werte Zufallsgeneratoren verwendet.

Bemerkenswert ist, dass man im dritten Schritt bei den deterministischen

matical form involving, however, unknown constants, called parameters. We call this representation a model, or more fully a probability model, for the data. 3. Given the form of the model, we regard the objective of the analysis to be the summarization of evidence about either the unknown parameters in the model or, occasionally, about the values of further random variables connected with the model, and, very importantly, the interpretation of that evidence.“ Bei diesen Ausführungen wird offenbar auf stochastische Datenmodelle Bezug genommen.

exogenen Variablen auf beliebige Weise vorgehen kann. Denn daraus folgt, dass man für die diesen Modellvariablen entsprechenden statistischen Variablen beliebige Verteilungen annehmen kann; und umgekehrt: dass das funktionale Modell diese Verteilungen unbestimmt lässt.

3. Funktionale Modelle für statistische Daten. Mithilfe eines funktionalen Modells können also statistische Daten mit ganz unterschiedlichen Verteilungen erzeugt werden. Umgekehrt gilt, dass zur Interpretation statistischer Daten stets unterschiedliche funktionale Modelle verwendet werden können.

Um das deutlich zu machen und einige damit verbundene Probleme zu besprechen, verwende ich ein Beispiel von N. Cartwright (1979). In diesem Beispiel gibt es eine Gesamtheit von Personen, ich nenne sie Ω, für die drei statistische Variablen definiert sind: X erfasst, ob es sich um einen Raucher ($X = 1$) handelt oder nicht ($X = 0$); Z erfasst, ob sich eine Person sportlich betätigt und gesundheitsfördernd ernährt ($Z = 1$) oder nicht ($Z = 0$); und Y erfasst, ob bei einer Person ein Herzinfarkt eingetreten ist ($Y = 1$) oder nicht ($Y = 0$). Es wird angenommen, dass folgende Häufigkeiten beobachtet worden sind:[4]

X	Z	$Y=0$	$Y=1$	insgesamt
0	0	420	80	500
0	1	95	5	100
1	0	70	30	100
1	1	450	50	500

(7.1)

Insoweit handelt es sich um einen statistischen Sachverhalt. Annahmen über kausale Zusammenhänge müssen sich demgegenüber auf ein funktionales Modell beziehen, das korrespondierende Modellvariablen (im Unterschied zu statistischen Variablen) verknüpft.

Eine mögliche Modellkonstruktion beginnt mit korrespondierend zu X, Z und Y definierten Modellvariablen $\ddot{X}$, $\ddot{Z}$ und $\dot{Y}$ und formuliert die Annahme, dass die Verteilung von $\dot{Y}$ von $\ddot{X}$ und $\ddot{Z}$ abhängig ist:

$$\begin{tikzcd} \ddot{X} \arrow[dr] & \\ & \dot{Y} \\ \ddot{Z} \arrow[ur] & \end{tikzcd} \qquad (7.2)$$

Dem Modell entspricht die stochastische Funktion

$$(x, z) \longrightarrow \Pr[\dot{Y} | \ddot{X} = x, \ddot{Z} = z]$$

Verwendet man die aus (7.1) berechenbaren bedingten Häufigkeiten als

[4] Diese Häufigkeiten sind willkürlich gewählt (so dass X und Z interaktive Bedingungen für Y bilden) und stammen nicht aus dem Beitrag von Cartwright.

Schätzwerte für die durch das Modell angenommenen bedingten Wahrscheinlichkeiten,[5] erhält man:

x	z	$\Pr(\dot{Y} = 0 \,\vert\, \ddot{X} = x, \ddot{Z} = z)$	$\Pr(\dot{Y} = 1 \,\vert\, \ddot{X} = x, \ddot{Z} = z)$
0	0	0.84	0.16
0	1	0.95	0.05
1	0	0.70	0.30
1	1	0.90	0.10

(7.3)

Innerhalb des Modells kann man nun eine Veränderung $\Delta(0, 1)$ bei der Variablen $\ddot{X}$ als eine mögliche funktionale Ursache für stochastische Effekte betrachten. Zur Illustration definieren wir einen stochastischen Effekt

$$\Delta^s(\dot{Y}; 0, 1; z) := \mathrm{E}(\dot{Y} | \ddot{X} = 1, \ddot{Z} = z) - \mathrm{E}(\dot{Y} | \ddot{X} = 0, \ddot{Z} = z)$$

durch den erfasst wird, wie sich im Kovariablenkontext $\ddot{Z} = z$ die Wahrscheinlichkeit eines Herzinfarkts zwischen Rauchern und Nichtrauchern unterscheidet. Mit den Werten in (7.3) findet man folgende kontextabhängigen Effekte:

$$\Delta^s(\dot{Y}; 0, 1; 0) = 0.30 - 0.16 = 0.14$$

$$\Delta^s(\dot{Y}; 0, 1; 1) = 0.10 - 0.05 = 0.05$$

Bei der Berechnung dieser Effekte wird vorausgesetzt, dass eine Veränderung $\Delta(0, 1)$ bei sich nicht ändernden Kovariablenkontexten möglich ist. Diese Voraussetzung wird unabhängig von den Daten durch das Modell angenommen, das von deterministischen Variablen $\ddot{X}$ und $\ddot{Z}$ ausgeht.

4. Unbestimmtheit durchschnittlicher Effekte. Die zuletzt genannte Voraussetzung ist vollständig unabhängig von der Verteilung der statistischen Variablen, die den deterministischen exogenen Modellvariablen entsprechen (also X und Z in unserem Beispiel). Wie in § 2 besprochen wurde, können für diese Variablen beliebige Verteilungen angenommen werden. Zu dem eben skizzierten Modell passen zum Beispiel auch folgende Daten:

X	Z	$Y=0$	$Y=1$	insgesamt
0	0	420	80	500
0	1	475	25	500
1	0	350	150	500
1	1	450	50	500

(7.4)

Man erhält die gleichen bedingten Wahrscheinlichkeiten, die in Tabelle (7.3) dargestellt worden sind, also auch die gleichen kontextabhängigen Effekte einer Veränderung $\Delta(0, 1)$ bei der Variablen $\ddot{X}$.

[5] D.h. $\mathrm{P}[Y | X = x, Z = z]$ dient als Schätzung für $\Pr[\dot{Y} | \ddot{X} = x, \ddot{Z} = z]$.

Es verändert sich jedoch der durchschnittliche Effekt, der von der Verteilung der Kovariablenkontexte in der jeweiligen Referenzgesamtheit abhängig ist. Geht man von den Daten in (7.1) aus, findet man:

$$\mathrm{P}(Y=1|X=1)-\mathrm{P}(Y=1|X=0)=\frac{80}{600}-\frac{85}{600}=-0.008$$

Geht man von den Daten in (7.4) aus, findet man:

$$\mathrm{P}(Y=1|X=1)-\mathrm{P}(Y=1|X=0)=\frac{200}{1000}-\frac{105}{1000}=0.095$$

Im ersten Fall ist die Häufigkeit eines Herzinfarkts bei Rauchern niedriger, im zweiten Fall ist sie größer als bei Nichtrauchern.

Es sollte jedoch betont werden, dass sich diese Berechnungen durchschnittlicher Effekte auf die jeweils verwendeten Daten beziehen; also nicht auf das funktionale Modell, sondern auf ein daraus gebildetes Datenmodell. Tatsächlich kann mit dem Modell (7.2) gar kein durchschnittlicher Effekt einer Veränderung $\Delta(x', x'')$ bei der Variablen $\ddot{X}$ berechnet werden, weil dieser Effekt kontextabhängig ist und der Kontext durch das Modell nicht bestimmt wird.[6]

5. Individual- und Populationsmodelle. Es sollte deutlich geworden sein, dass es keinen unmittelbaren und eindeutigen Zusammenhang zwischen funktionalen Modellen und statistischen Daten gibt. Das wird verständlich, wenn man berücksichtigt, dass sich funktionale Modelle (so wie sie bisher besprochen wurden) gar nicht auf statistische Gesamtheiten beziehen, sondern dass es sich um generische Individualmodelle handelt, die sich in unbestimmter Weise auf Individuen (individuelle Objekte oder Situationen) beziehen. Zwar kann es sich dabei auch um die Elemente einer statistischen Gesamtheit handeln. Aber die Aufgabe eines funktionalen Modells besteht nicht darin, die Merkmalsverteilungen in irgendeiner bestimmten statistischen Gesamtheit darzustellen.

Bezieht man sich dagegen auf die Frage, wie statistische Verteilungen zustande kommen und sich verändern, sind generische Individualmodelle unzureichend, stattdessen müssen Modelle verwendet werden, die es erlauben, statistische Verteilungen als Werte von Modellvariablen zu betrachten. Mit solchen Modellen, die wir *Populationsmodelle* nennen, beschäftigen wir uns in Kapitel **??**.

[6]Dieser Sachverhalt wird natürlich nicht sichtbar, wenn von vornherein abgeschlossene funktionale Modelle verwendet werden, bei denen für alle exogenen Variablen eine gegebene Wahrscheinlichkeitsverteilung angenommen wird. Man vgl. beispielsweise den Modellansatz bei Woodward (2001: 41).

7.2 Experimente und Beobachtungsdaten

Passives Beobachten und aktives Experimentieren zu unterscheiden, hat eine lange Tradition.[7] In diesem Abschnitt beschäftigen wir uns mit der Unterscheidung im Hinblick auf Möglichkeiten, funktionale Modelle mit Daten zu verknüpfen. Wiederum geht es nur um konzeptionelle Fragen, nicht darum, wie man parametrische Modelle mit Daten schätzen kann. Zunächst beziehen wir uns auf randomisierte Experimente, die zur Begründung kausal interpretierbarer Modelle oft als besonders geeignet angesehen werden. Dann soll gezeigt werden, dass diese Idee bei sozialwissenschaftlichen Anwendungen auf Sinngrenzen stößt.

1. Randomisierte Experimente. Wir beziehen uns auf ein einfaches funktionales Modell der Form

$$\ddot{X} \longrightarrow\!\!\!\!\!\rightarrow \dot{Y} \tag{7.5}$$

das mit experimentell erzeugbaren Daten geschätzt werden soll. Daten können für eine Gesamtheit von Untersuchungseinheiten (Personen, Objekte oder Situationen) erzeugt werden, die wir wie bisher durch $\Omega = \{\omega_1, \ldots, \omega_n\}$ bezeichnen. Ein Experimentator kann für jede Untersuchungseinheit einen Wert der Variablen $\ddot{X}$ festlegen und dann beobachten, welcher Wert der Variablen $\dot{Y}$ angenommen wird. Auf diese Weise entstehen Werte einer zweidimensionalen statistischen Variablen

$$(X, Y): \Omega \longrightarrow \tilde{\mathcal{X}} \times \tilde{\mathcal{Y}} \tag{7.6}$$

die verwendet werden können, um (eine Modellfunktion für) die Regressionsfunktion $x \longrightarrow \mathrm{P}[Y|X=x]$ (oder irgendeine daraus ableitbare spezielle Regressionsfunktion) zu berechnen. Diese Funktion kann dann als eine Schätzung der durch das Modell (7.5) angenommenen stochastischen Funktion $x \longrightarrow \Pr[\dot{Y}|\ddot{X}=x]$ angesehen werden.

Von einem *randomisierten Experiment* spricht man, wenn bei der Zuordnung der Werte von $\ddot{X}$ zu den Untersuchungseinheiten in Ω (also bei der Erzeugung von Werten der statistischen Variablen X) ein Zufallsgenerator mit einer bekannten Verteilung verwendet wird.[8] Warum erscheint

[7]Der Philosoph d'Alembert hat die Unterscheidung in der zusammen mit Diderot herausgegebenen „Enzyklopädie“ (1756/1984: 434) so formuliert: „Die Beobachtung, die weniger originell und tiefgründig ist, beschränkt sich auf die Tatsachen, die man vor Augen hat, das heißt darauf, Erscheinungen aller Art, die uns das Schauspiel der Natur darbietet, gut zu betrachten und ausführlich zu beschreiben; dagegen sucht das Experiment die Natur tiefer zu erforschen, ihr das zu entreißen, was sie verbirgt, und durch mannigfache Kombination der Körper neue Erscheinungen hervorzubringen, um diese wiederum zu studieren – kurz, es beschränkt sich nicht darauf, die Natur zu belauschen, sondern es befragt sie und zwingt sie zur Auskunft.“

[8]In einer Formulierung von Holland (1994: 269): „Randomization is a physical act in which a known chance mechanism is used in particular ways to construct the function x

eine solche Randomisierung vorteilhaft? Man kann sich vorstellen, dass $\dot{Y}$ nicht nur von $\ddot{X}$, sondern auch von Werten einer weiteren Variablen $\ddot{Z}$ abhängt, also ein Modell der Form

$$\ddot{X} \twoheadrightarrow \dot{Y} \twoheadleftarrow \ddot{Z} \qquad (7.7)$$

annehmen. Wäre $\ddot{Z}$ beobachtbar, könnte man anstelle von (7.6) Werte einer statistischen Variablen

$$(X, Y, Z) : \Omega \longrightarrow \tilde{\mathcal{X}} \times \tilde{\mathcal{Y}} \times \tilde{\mathcal{Z}} \qquad (7.8)$$

verwenden und untersuchen, ob und ggf. wie Effekte von Veränderungen $\Delta(x', x'')$ von Werten von $\ddot{Z}$ abhängig sind. Wenn Werte von $\ddot{Z}$ nicht beobachtet werden können, ist das nicht möglich, sondern man muss sich an einem Modell der Form

$$\ddot{X} \twoheadrightarrow \dot{Y} \twoheadleftarrow \dot{Z} \qquad (7.9)$$

orientieren, so dass (7.5) als ein reduziertes Modell betrachtet werden kann (vgl. Abschnitt 5.3). Dann entstehen zwei Fragen: Ist die für (7.9) vorausgesetzte Annahme richtig, dass $\dot{Z}$ von $\ddot{X}$ stochastisch unabhängig ist? Welche Verteilung kann für $\dot{Z}$ angenommen werden?

Randomisierung bezieht sich auf die erste Frage und lässt es plausibel erscheinen, die aus einem randomisierten Experiment gewonnenen Daten durch ein Modell der Form (7.9) zu interpretieren.[9] Vollständig unabhängig von der Randomisierung liefert die empirische Bezugnahme auf eine bestimmte Gesamtheit Ω auch einen Zugang zur zweiten Frage. Man kann annehmen (genauer: festlegen[10]), dass die Verteilung von $\dot{Z}$ näherungsweise mit der statistischen Verteilung von Z in Ω übereinstimmt:

$$\Pr[\dot{Z}] \approx \mathrm{P}[Z] \qquad (7.10)$$

Zwar kann man Z nicht beobachten; die Annahme erlaubt aber eine Antwort auf die Frage, worauf sich die Durchschnittsbildung bezieht, wenn ein Modell der Form (7.9) zur Interpretation des ursprünglichen Modells (7.5) verwendet wird, nämlich: Die mithilfe der Daten (7.6) berechenbaren Effekte einer Veränderung $\Delta(x', x'')$ entsprechen einer Verteilung von Kovariablenkontexten in der experimentellen Gesamtheit Ω.

[womit in unserer Notation die statistische Variable X gemeint ist].“ Eine ausführliche Diskussion unterschiedlicher Methoden findet man bei Shadish, Cook und Campbell (2002).

[9] Hier könnte man auch so argumentieren: Wenn man Werte von (X, Z) beobachten könnte, wären infolge der Randomisierung X und Z näherungsweise statistisch unabhängig.

[10] Denn die Festlegung einer Verteilung für die Modellvariable $\dot{Z}$ bildet einen Teil der Definition des Modells (7.9).

2. Was wird durch Randomisierung erreicht? Randomisierung kann also als eine Methode zur Konstruktion durchschnittlicher Effekte angesehen werden. Es ist aber wichtig zu verstehen, dass man gleichwohl von den Verteilungen der unbeobachteten Variablen in der jeweiligen experimentellen Gesamtheit abhängig bleibt.

Um diesen Sachverhalt zu illustrieren, betrachten wir eine Variante des Beispiels aus Abschnitt 6.3 (§ 6). $\ddot{X}$ erfasst den Schultyp, $\dot{Y}$ den schulischen Erfolg, beide Variablen können die Werte 0 oder 1 annehmen. Außerdem gibt es eine unbeobachtete Variable $\ddot{Z}$, die das Bildungsniveau der Eltern erfasst (0 niedrig, 1 hoch).[11] Für das Modell (7.7) wird wie in Abschnitt 6.3 folgende stochastische Funktion angenommen:

z	x	$\mathrm{E}(\dot{Y} \mid \ddot{X} = x, \ddot{Z} = z)$
0	0	0.5
0	1	0.7
1	0	0.8
1	1	0.9

(7.11)

Als Voraussetzung für die Durchführbarkeit eines randomisierten Experiments wird angenommen, dass die Schulkinder unabhängig vom Bildungsniveau ihrer Eltern auf die beiden Schultypen verteilt werden können.

Nehmen wir nun an, dass diese Voraussetzung, um eine Randomisierung vornehmen zu können, gegeben ist. Jedem Schüler wird mit der Wahrscheinlichkeit p_x der Schultyp 1, mit der Wahrscheinlichkeit $1 - p_x$ der Schultyp 0 zugewiesen, so dass eine statistische Variable $X : \Omega \longrightarrow \{0, 1\}$ mit der Verteilung $\mathrm{P}(X = 1) \approx p_x$ entsteht. Schließlich kann irgendwann der schulische Erfolg ermittelt werden, so dass man für jeden Schüler ω sowohl den Schultyp $X(\omega)$ als auch den schulischen Erfolg $Y(\omega)$ kennt. Außerdem unterstellen wir die Existenz einer statistischen Variablen Z, die die nicht-beobachteten Bildungsniveaus der Eltern erfasst; $p_z := \mathrm{P}(Z = 1)$ sei der Anteil der Schüler in Ω, deren Eltern ein hohes Bildungsniveau haben.

Zu vergleichen sind die Mittelwerte $\mathrm{M}(Y \mid X = 1)$ und $\mathrm{M}(Y \mid X = 0)$. Durch das Randomisierungsverfahren wird erreicht, dass X und Z näherungsweise statistisch unabhängig sind, so dass man (wenn man sich auf das Modell (7.9) bezieht) zur Berechnung folgende Näherung verwenden kann:

$$\mathrm{M}(Y \mid X = x) \approx \mathrm{E}(\dot{Y} \mid \ddot{X} = x, \dot{Z} = 1)\, p_z + \mathrm{E}(\dot{Y} \mid \ddot{X} = x, \dot{Z} = 0)\, (1 - p_z)$$

Verwendet man die in (7.11) angegebenen Werte, findet man:

$$\mathrm{M}(Y \mid X = 1) - \mathrm{M}(Y \mid X = 0) \approx 0.2 - 0.1\, p_z$$

[11] Um den Notationen dieses Abschnitts zu entsprechen, werden die Variablennamen für den Schultyp und das Bildungsniveau der Eltern vertauscht.

Was also ist „der“ Effekt des Schultyps für den schulischen Erfolg? Die Antwort, die das randomisierte Experiment liefert, hängt offenbar auch von der Verteilung der Bildungsniveaus der Eltern in der für das Experiment verwendeten Gesamtheit von Schülern ab.

3. Das Standardargument für Randomisierung. Ein oft angeführtes Argument besteht darin, dass es durch Randomisierung möglich wird, den Effekt einer Veränderung des Werts einer Variablen $\ddot{X}$ durch einen Vergleich von Gruppen von Untersuchungseinheiten zu ermitteln, bei denen alle übrigen Variablen eine ähnliche Verteilung aufweisen:

„The key virtue of randomization is to create balanced treatment and control groups that resemble each other across all causally relevant variables except treatment status.“ (Elwert und Winship 2002: 433)

Auf den ersten Blick erscheint die Überlegung plausibel. Um den Effekt einer Veränderung $\Delta(x_0, x_1)$ bei einer Variablen $\ddot{X}$ für eine abhängige Variable $\dot{Y}$ zu ermitteln, muss man sich auf eine korrespondierend zu $\ddot{X}$ und $\dot{Y}$ definierte statistische Variable (X, Y) mit einer Bezugsgesamtheit Ω beziehen und zwei Gruppen vergleichen:

$$\Omega_1 := \{\omega \,|\, X(\omega) = x_1\} \quad \text{und} \quad \Omega_0 := \{\omega \,|\, X(\omega) = x_0\}$$

Bezieht man sich auf Experimente, ist Ω_1 die Treatment- und Ω_0 die Kontrollgruppe. Es erscheint plausibel: Wenn die Werte von X durch ein Randomisierungsverfahren entstehen, sind die beiden Gruppen im Hinblick auf alle bereits vor der Randomisierung fixierten Merkmale ähnlich. Aber warum ist das für die Interpretation des schließlich ermittelten Effekts von $\Delta(x_0, x_1)$ relevant?

Denken wir an irgendeine Variable Z, bei der sich Ω_0 und Ω_1 unterscheiden könnten. Es gibt zwei Möglichkeiten. Entweder ist Z für den Effekt irrelevant; dann ist eine Randomisierung bzgl. dieser Variablen nicht erforderlich. Oder man kann vermuten, dass der zu ermittelnde Effekt auch von Werten von Z abhängt. Betrachten wir also diese Möglichkeit.

Wenn die experimentelle Gesamtheit Ω hinreichend groß ist, kann man annehmen, dass infolge der Randomisierung X und Z näherungsweise unabhängig sind: $\mathrm{P}[Z|X = x_0] \approx \mathrm{P}[Z|X = x_1]$. Infolgedessen kann der Effekt von $\Delta(x_0, x_1)$ als ein einfacher Durchschnitt dargestellt werden:

$$\mathrm{M}(Y|X = x_1) - \mathrm{M}(Y|X = x_0) \approx \Delta^a(Y|x_0, x_1; Z) \qquad (7.12)$$

mit der Definition

$$\Delta^a(Y|x_0, x_1; Z) := \qquad (7.13)$$
$$\textstyle\sum_z \left(\mathrm{M}(Y|X = x_1, Z = z) - \mathrm{M}(Y|X = x_0, Z = z)\right) \mathrm{P}(Z = z)$$

Aber daraus folgt nicht, dass der mit den Daten ermittelte durchschnittliche Effekt unabhängig von Z ist und nur der Ursache $\Delta(x_0, x_1)$ zugerechnet werden kann. Wie in § 2 anhand eines Beispiels gezeigt wurde, kann

der Effekt auch noch von der Verteilung von Z in der experimentellen Gesamtheit abhängen. Über mögliche Effekte in einer anderen Gesamtheit mit einer anderen Verteilung von Z lassen sich keine Aussagen machen. Dafür wäre es erforderlich, die kontextspezifischen Effekte

$$\mathrm{M}(Y|X = x_1, Z = z) - \mathrm{M}(Y|X = x_0, Z = z)$$

zu ermitteln, also zunächst Daten für Z zu gewinnen.

4. Unterschiedliche durchschnittliche Effekte. Die Überlegung kann fortgesetzt werden: Könnte man Daten für Z gewinnen, wäre es nicht erforderlich, bzgl. dieser Variablen eine Randomisierung vorzunehmen, denn man könnte dann kontextspezifische Effekte schätzen. Zwar wäre ohne eine Randomisierung nicht gewährleistet, dass X und Z näherungsweise unabhängig sind, und infolgedessen könnte (7.12) falsch sein. Aber wieder stellt sich die Frage, warum das für die Definition eines kausalen Effekts relevant sein sollte.

Das Problem resultiert daraus, dass man einerseits einen bestimmten kausalen Effekt schätzen möchte, andererseits aber weiß oder annimmt, dass der Effekt kontextabhängig ist. In unserem Beispiel hängt der Effekt einer Ursache $\Delta(x_0, x_1)$ auch von den Werten von Z ab. Um dieser Ursache einen bestimmten Effekt zuzurechnen, muss also entweder auf einen bestimmten Kovariablenkontext $Z = z$ Bezug genommen werden; oder es muss ein Durchschnitt gebildet werden. So betrachtet ist (7.13) ein Vorschlag, wie eine Durchschnittsbildung vorgenommen werden sollte; aber es gibt beliebig viele andere Möglichkeiten.

Was spricht für die in (7.13) vorgeschlagene Durchschnittsbildung? Eine mögliche Überlegung bezieht sich auf ein „zufällig“ aus Ω ausgewähltes Individuum ω. Wenn man den Kovariablenkontext $Z(\omega)$ nicht kennt, sondern nur weiß, dass ω einer Gesamtheit mit der Verteilung $\mathrm{P}[Z]$ angehört, erscheint es plausibel, zur Berechnung eines Erwartungswerts die durch (7.13) vorgeschlagene Durchschnittsbildung zu verwenden.

Ebensogut kann man sich andere Anwendungskontexte vorstellen. Zur Illustration verwenden wir wieder das Beispiel aus § 2 und nehmen an, dass 40 % der Schüler in der Referenzgesamtheit Ω Eltern mit einem höheren Bildungsniveau haben und somit $\mathrm{P}(Z = 1) = 0.4$ ist. Außerdem nehmen wir an, dass die Zuordnung eines Schultyps auf folgende Weise vom Bildungsniveau der Eltern abhängt:

$$\mathrm{P}(X = 1|Z = 0) = 0.4 \quad \text{und} \quad \mathrm{P}(X = 1|Z = 1) = 0.8$$

Mit den Daten aus Tabelle (7.11) findet man für den in (7.13) definierten durchschnittlichen Effekt: $(0.7 - 0.5)\, 0.6 + (0.9 - 0.8)\, 0.4 = 0.16$. Bezieht man sich auf ein „zufällig“ aus Ω ausgewähltes Schulkind, könnte man also erwarten, dass die Wahrscheinlichkeit eines erfolgreichen Schulabschlusses beim Schultyp 1 um 0.16 größer ist als beim Schultyp 0.

Aber man kann auch eine andere und vielleicht interessantere Frage stellen: Wie würde sich die Wahrscheinlichkeit eines erfolgreichen Schulabschlusses verändern, wenn ein Schulkind aus dem Schultyp 0 in den Schultyp 1 wechselt? Dann wird ein Schulkind „zufällig“ aus der Teilgesamtheit $\Omega_0 = \{\omega | X(\omega) = 0\}$ ausgewählt, in der die Bildungsniveaus der Eltern folgendermaßen verteilt sind: $\mathrm{P}(Z = 0 \,|\, X = 0) = 0.82$ und $\mathrm{P}(Z = 1 \,|\, X = 0) = 0.18$, und man findet einen größeren durchschnittlichen Effekt, nämlich $(0.7 - 0.5)\, 0.82 + (0.9 - 0.8)\, 0.18 = 0.18$.

5. Argumente mit potentiellen Individualeffekten. Eine Variante des Standardarguments für Randomisierung stammt von Autoren, die mit potentiellen Individualeffekten („potential outcomes“) argumentieren. Wir knüpfen hier an Hollands Darstellung dieses Ansatzes an, die in Abschnitt 6.2 besprochen wurde. Mit den dort eingeführten Notationen kann man den Gedankengang anhand folgender Tabelle erklären:

ω	$X(\omega)$	$Y_0(\omega)$	$Y_1(\omega)$	$Y(\omega)$
ω_1	x_1	y_{01}	y_{11}	y_1
ω_2	x_2	y_{02}	y_{12}	y_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
ω_n	x_n	y_{0n}	y_{1n}	y_n

(7.14)

Die Tabelle bezieht sich auf die Untersuchungseinheiten in der experimentellen Gesamtheit Ω. Jede Untersuchungseinheit ω_i kann entweder einem Einfluss $X = 0$ oder $X = 1$ ausgesetzt werden. $Y_x(\omega_i)$ ist das Ergebnis, das resultieren würde, wenn ω_i dem Einfluss $X = x$ ausgesetzt würde.[12] $X(\omega_i)$ und $Y(\omega_i)$ erfassen, welchem Einfluss ω_i tatsächlich ausgesetzt wurde und welches Resultat daraus folgte:

$$Y(\omega_i) = \begin{cases} Y_0(\omega_i) & \text{wenn } X(\omega_i) = 0 \\ Y_1(\omega_i) & \text{wenn } X(\omega_i) = 1 \end{cases}$$

Das Experiment soll dazu dienen, einen durchschnittlichen Effekt

$$\mathrm{M}(Y_1 - Y_0) = \textstyle\sum_{i=1,n} (Y_1(\omega_i) - Y_0(\omega_i))/n$$

zu ermitteln. Da jede Untersuchungseinheit nur einem der beiden Einflüsse ausgesetzt werden kann, muss überlegt werden, wie das möglich ist. Das Argument lautet,[13] Werte von X durch ein Randomisierungsverfahren zu erzeugen; denn dann sind $\Omega_0 = \{\omega \,|\, X(\omega) = 0\}$ und $\Omega_1 = \{\omega \,|\, X(\omega) = 1\}$ einfache Zufallsstichproben aus Ω und man erhält unverzerrte Schätzwerte $\mathrm{M}(Y_0) \approx \mathrm{M}(Y|X = 0)$ und $\mathrm{M}(Y_1) \approx \mathrm{M}(Y|X = 1)$, also

$$\mathrm{M}(Y_1 - Y_0) \approx \mathrm{M}(Y|X = 1) - \mathrm{M}(Y|X = 0)$$

[12] Dabei wird angenommen, dass Werte dieser Variablen unabhängig davon existieren, welchen Einflüssen die Individuen tatsächlich ausgesetzt werden. Diese Annahme muss offenbar vorausgesetzt werden, um die Tabelle (7.14) formulieren zu können.

[13] Man vgl. beispielsweise Holland (1986: 948-9).

Die entscheidende Voraussetzung dieses Gedankengangs besteht in der Annahme, dass es Werte von Y_0 und Y_1 gibt, die sich den Untersuchungseinheiten in Ω bereits vor der Festlegung von Werten für X als konstante Größen zurechnen lassen. Diese Annahme ermöglicht es, Randomisierung so zu betrachten, als ob es sich um eine Erzeugung von Zufallsstichproben handelt.

6. Kausale und pseudo-deskriptive Fragestellungen. Man kann sich den theoretischen Ansatz anhand eines Urnenmodells verdeutlichen.[14] Für jedes ω_i gibt es in der Urne eine Kugel, auf deren Vorderseite y_{0i} und auf deren Rückseite y_{1i} eingetragen ist. Dann werden zwei Zufallsstichproben gezogen und bei der ersten die Werte der Vorderseite, bei der zweiten die Werte der Rückseite festgestellt.

Eine solche Betrachtungsweise widerspricht jedoch dem Sinn einer kausalen Fragestellung, für die die Annahme eines substantiellen Prozesses, der von einer Ursache zu einer Wirkung führt, konstitutiv ist. Wie ein solcher Prozess konzipiert werden kann, hängt auch davon ab, ob man sich auf dynamische oder auf komparative Ursachen bezieht. Aber die Idee kann bereits angedeutet werden, wenn man den Wert einer Variablen X nur als Bedingung für den Wert einer Variablen Y annimmt:

$$X(\omega_i) \xrightarrow{\text{kausaler Prozess}} Y(\omega_i)$$

An die Stelle eines solchen kausalen Prozesses tritt bei der Argumentation mit potentiellen Effekten ein Selektionsprozess, für dessen Konzeptualisierung angenommen werden muss, dass die möglichen Ergebnisse bereits vor dem Beginn des Selektionsprozesses existieren.

Zur Verdeutlichung kann wieder unser Schulbeispiel dienen. Durch $X(\omega_i)$ wird der Schultyp festgelegt, der dann eine Bedingung für einen sich anschließenden Prozess bildet, in dem sich der schulische Erfolg $Y(\omega_i)$ herausstellt. Offenbar ist es wichtig, dass man sich einen solchen Prozess vorstellen kann, für den $X(\omega_i)$ eine relevante Bedingung bildet.

Das ist insbesondere dann wichtig, wenn mit der Möglichkeit weiterer Ursachen (Bedingungen, die unterschiedlich beschaffen sein können) gerechnet werden muss, die im Anschluss an das Entstehen der Bedingung $X(\omega_i)$ für das Ergebnis relevant sein können. Um bei unserem Beispiel zu bleiben, nehmen wir an, dass der schulische Erfolg auch davon abhängt, ob sich die Eltern viel $(X' = 1)$ oder wenig $(X' = 0)$ Zeit nehmen, dem Kind bei seinen Schularbeiten zu helfen. Folgte man dem Ansatz der potentiellen Individualeffekte, könnten dann zusätzlich zu den fiktiven Individualvariablen Y_x weitere gleichermaßen fiktive Variablen $Y'_{x'}$ definiert werden, die

[14] Diesen Vorschlag hat bereits Neyman gemacht, der oft als einer der Begründer von Argumenten mit potentiellen Effekten angesehen wird; vgl. Splawa-Neyman (1990) und den Kommentar von Rubin (1990).

den Schulerfolg angeben, wenn $X' = x'$ ist. Als Erweiterung von (7.14) würde folgendes Schema entstehen:

ω	$X(\omega)$	$X'(\omega)$	$Y_0(\omega)$	$Y_1(\omega)$	$Y_0'(\omega)$	$Y_1'(\omega)$
ω_1	x_1	x_1'	y_{01}	y_{11}	y_{01}'	y_{11}'
ω_2	x_2	x_2'	y_{02}	y_{12}	y_{02}'	y_{12}'
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
ω_n	x_n	x_n'	y_{0n}	y_{1n}	y_{0n}'	y_{1n}'

Man erkennt, dass das Schema die Möglichkeit ausschließt, X und X' unabhängig voneinander zu variieren. Denn wenn bei einem Schulkind ω_i der Schultyp x_i festgelegt worden ist, ist dadurch bereits sein Schulerfolg determiniert: $Y_{x_i}(\omega_i)$; und zwar unabhängig davon, welchen Wert $X'(\omega_i)$ annimmt.[15] Man kann sich beispielsweise folgende Werte vorstellen:

$$y_{0i} = 0,\ \ y_{1i} = 1,\ \ y_{0i}' = 0,\ \ y_{1i}' = 1$$

Wäre dann $x_i = 1$ und $x_i' = 0$, wäre auch der schulische Erfolg gleichzeitig 1 und 0. Ein Ansatz, der mit potentiellen Individualeffekten beginnt, eignet sich also nicht, wenn ein Effekt gleichzeitig von zwei (oder mehr) Ursachen abhängig sein kann.[16]

Man kann sagen, dass bei der Argumentation mit potentiellen Individualeffekten an die Stelle einer kausalen Fragestellung, die einen kausalen Prozess betrifft, eine pseudo-deskriptive Fragestellung tritt. Eine solche Fragestellung erscheint sinnvoll möglich, wenn und insoweit man sich auf Daten bezieht, also auf bereits realisierte Ursachen und Wirkungen. Dann kann man eine kontrafaktische Frage stellen: Welche Werte der Variablen Y wären bei den Mitgliedern der jeweiligen Datengesamtheit Ω entstanden, wenn sie nicht den tatsächlich realisierten Einflüssen $X(\omega)$, sondern anderen Einflüssen ausgesetzt worden wären? Aber selbst wenn man diese Frage beantworten könnte, wäre die Antwort nur für die Datengesamtheit Ω relevant und hätte keinerlei Implikationen für kausale Fragestellungen, die sich auf andere Individuen oder Situationen beziehen.

7. Sinngrenzen der Randomisierung. Bei Experimenten kann ein Experimentator auf die Bedingungen, unter denen die Werte einer abhängigen Variablen entstehen, Einfluss nehmen. Von *Beobachtungsdaten* wird

[15] Es sei angemerkt, dass das Problem nicht daraus resultiert, dass in dem Ansatz von Holland, dem wir hier folgen, die potentiellen Effekte als deterministische Größen konzipiert werden. Die Probleme und Inkonsistenzen, auf die hier hingewiesen wird, entstehen ebenso, wenn die fiktiven Variablen Y_x und $Y_{x'}'$ als stochastische Größen aufgefasst werden.

[16] Zwar könnte man die unterschiedlichen Ursachen kombinieren; in unserem Beispiel könnte man die Variablen X und X' zu einer Variablen X'' zusammenfassen und für jeden ihrer vier möglichen Werte einen bestimmten individuellen Schulerfolg annehmen. Dann aber wäre es nicht mehr möglich, für die beteiligten Variablen (X und X') jeweils separat bestimmte kausale Effekte anzunehmen; und insbesondere wäre es nicht möglich, bezüglich X und X' simultan zu randomisieren.

gesprochen, wenn dies nicht möglich ist, also die Werte aller jeweils relevanten Variablen (bestenfalls) passiv beobachtet werden können. Insbesondere können also Beobachtungsdaten nicht als Ergebnis eines randomisierten Experiments angesehen werden. Autoren, die kausale Effekte zunächst durch randomisierte Experimente definieren, sehen darin oft ein Problem.[17] So schreibt beispielsweise Rosenbaum (1984: 41):

„The purpose of an observational study is to “elucidate cause-and-effect relationships.” An assessment of the evidence concerning the extent to which the treatment actually causes its apparent effects is, therefore, central and necessary. There are, however, difficulties involved. The most familiar difficulty is that, since treatments were not randomly assigned to experimental units, the treated and control groups may not be directly comparable.“

Beobachtungsdaten können für unterschiedliche, auch primär deskriptive Zwecke verwendet werden. Hier interessiert, wie sie für kausale Fragestellungen verwendet werden können.[18] Aber was ist damit gemeint?

Eine Variante der Fragestellung ist: Wie kann man mithilfe von Beobachtungsdaten kausale Effekte schätzen, die durch eine vorgängige Orientierung an randomisierten Experimenten definiert sind. Die leitende Idee besteht dann darin, die Beobachtungsdaten aus der Perspektive eines vergleichbaren randomisierten Experiments zu betrachten.[19] Im Hinblick auf sozialwissenschaftliche Anwendungen ist es jedoch wichtig, sich nicht von vornherein auf diese Variante der Fragestellung festzulegen.

Um die Überlegung zu erläutern, sei angenommen, dass man sich für ein funktionales Modell $\ddot{X} \longrightarrow\!\!\!\!\!\rightarrow \dot{Y}$ interessiert, zum Beispiel für einen Zusammenhang zwischen Schultyp ($\ddot{X}$) und Schulerfolg ($\dot{Y}$). Bei der Formulierung des Modells wird offen gelassen, wie Werte von $\ddot{X}$ entstehen. Man kann sich an einem Bild der folgenden Art orientieren:

$$\begin{array}{ccc} \boxed{\text{Substantieller Prozess}} & \longrightarrow & \boxed{\text{Werte von } \ddot{X}} \\ & & \vdots \\ & & \boxed{\text{Beobachtungen}} \end{array} \qquad (7.15)$$

Zunächst gibt es einen substantiellen Prozess, durch den die als Werte von $\ddot{X}$ fixierbaren Sachverhalte entstehen, dann können durch Beobachtungen dieser Sachverhalte Daten entstehen. In unserem Beispiel werden durch den substantiellen Prozess Schüler auf Schultypen verteilt; hinterher kann ein datenerzeugender Prozess Beobachtungen der realisierten Zuordnungen liefern. Diese Überlegung macht auch deutlich, dass Randomisierung selbst

[17] Zum Beispiel: Collier, Brady und Seawright (2004: 230-3).

[18] Wenn in der englischsprachigen Literatur von „observational studies“ gesprochen wird, ist meistens – wie in dem oben angeführten Zitat von Rosenbaum – gemeint, dass man sich für kausale Fragestellungen interessiert.

[19] Diesen Grundgedanken verfolgt beispielsweise Rosenbaum (2002).

eine Variante eines substantiellen Prozesses ist.

Betrachten wir jetzt die Frage, wie man einer Veränderung $\Delta(x', x'')$ einen bestimmten kausalen Effekt zurechnen kann. Es scheint so, dass es dafür irrelevant ist, wie eine solche Ursache entsteht oder entstanden ist.[20] Das ist jedoch nur dann richtig, wenn die Entstehung der Ursache nicht in einer für ihren Effekt relevanten Weise kontextabhängig ist. Um das zu verdeutlichen, bleiben wir bei unserem Beispiel und nehmen an, dass der schulische Erfolg nicht nur vom Schultyp, sondern auch vom Bildungsniveau der Eltern ($\ddot{Z}$) abhängt. Dann kann man folgende Möglichkeiten betrachten:

$$
\text{(a)} \quad \begin{tikzcd}[row sep=small] \ddot{Z} \arrow[dr] & \\ & \dot{Y} \\ \ddot{X} \arrow[ur] & \end{tikzcd}
\qquad\qquad
\text{(b)} \quad \begin{tikzcd}[row sep=small] \ddot{Z} \arrow[dr] \arrow[dd] & \\ & \dot{Y} \\ \dot{X} \arrow[ur] & \end{tikzcd}
\tag{7.16}
$$

Im Modell (a) wird nicht näher angegeben, wie Werte von $\ddot{X}$ entstehen; im Modell (b) wird angenommen, dass der Prozess, durch den Schülern Schultypen zugewiesen werden, auch vom Bildungsniveau ihrer Eltern abhängt. Das Modell (b) ist infolgedessen mit der Vorstellung einer Randomisierung, die bei der Variablen $\dot{X}$ ansetzt, nicht vereinbar; und deshalb liefert in diesem Fall auch der Verweis auf eine tatsächlich oder fiktiv durchgeführte Randomisierung keine angemessene Definition des kausalen Effekts einer Veränderung $\Delta(x', x'')$.

Das Problem besteht nicht in erster Linie darin, dass eine Randomisierung oft nicht möglich ist, vielmehr darin, dass eine Randomisierung mit den Prozessen, durch die in der sozialen Realität Ursachen entstehen, nicht vereinbar sein kann.[21] In unserem Beispiel: Selbst wenn es möglich wäre, im Rahmen eines Experiments Schüler zufällig auf Schultypen zu verteilen, wäre das Ergebnis eines solchen Experiments nicht relevant, weil der in seinen kausalen Aspekten aufzuklärende Prozess dem Modell widerspricht, an dem sich das randomisierte Experiment orientiert.

Definitionen kausaler Effekte, die sich an randomisierten Experimenten orientieren, können also nicht ohne weiteres als ein Ideal angesehen werden. Vielmehr ist umgekehrt die Frage zu stellen, für welche Anwendungskontexte ein aus solchen Experimenten gewinnbares Kausalwissen relevant sein kann. Die Frage ist insbesondere dann wichtig, wenn das intendierte Kausalwissen dem Zweck dienen soll, Folgen von Handlungsalternativen einschätzbar zu machen; denn dann muss bedacht werden, welche Akteure die Entscheidungen (zum Beispiel über die Auswahl eines Schultyps), auf die sich die Theoriebildung bezieht, tatsächlich treffen.

[20] Konzeptionalisierungen der Prozesse, durch die Ursachen entstehen, müssen insbesondere berücksichtigen, ob es sich um dynamische oder komparative Ursachen handelt. An dieser Stelle genügt es jedoch, die Frage, wie eine Ursache $\Delta(x', x'')$ entsteht, in die unspezifische Frage zu übersetzen, wie die Werte x' und x'' entstehen.

[21] Man vgl. hierzu auch Heckman (1992).

7.3 Prozesse und Bezugsprobleme

Im vorangegangenen Abschnitt wurde zu zeigen versucht, dass eine Orientierung an randomisierten Experimenten an Sinngrenzen stößt, wenn man sich mit funktionalen Modellen auf soziale Prozesse beziehen möchte. Einige weitere Konsequenzen, die daraus folgen, dass an sozialen Prozessen Akteure mit eigenen Vorstellungen beteiligt sind, werden in diesem Abschnitt besprochen.

1. Orientierung an Interventionen. Dass Kausalwissen oft dem Zweck dienen soll, Folgen von Interventionen (womit hier absichtlich erzeugte Ursachen gemeint sind) einschätzbar zu machen, ist offenkundig. Einige Autoren haben versucht, hieraus einen Leitgedanken zur Definition kausaler Beziehungen zu gewinnen.[22] Die Idee liegt nahe, wenn man sich auf Experimente bezieht, bei denen mögliche Ursachen absichtlich erzeugt werden können; bei sozialwissenschaftlichen Anwendungen muss jedoch bedacht werden, dass es bereits im Gegenstandsbereich der Modellbildung Akteure gibt – als Träger der zu modellierenden substantiellen Prozesse – und dass es zunächst diese Akteure sind, die Interventionen vornehmen und dadurch Effekte bewirken können.

Bevor das näher ausgeführt wird, soll darauf hingewiesen werden, dass sich die Vorstellung einer (fiktiven) Intervention zunächst nur für deterministische exogene Modellvariablen eignet. Diese Variablen sind gerade so definiert, dass man für sie unterschiedliche Werte annehmen kann. Anders verhält es sich bei endogenen Modellvariablen, deren Werte von anderen Modellvariablen abhängen, so dass sie nicht beliebig verändert werden können. Zur Verdeutlichung beziehen wir uns auf die in (7.16) angegebenen Varianten für unser Schulbeispiel. In der Modellvariante (a) wird der Schultyp durch eine exogene Variable $\ddot{X}$ repräsentiert, so dass im Rahmen des Modells beliebige Annahmen möglich sind. In der Modellvariante (b) gibt es stattdessen eine endogene Variable $\dot{X}$, deren Werte von $\ddot{Z}$, dem Bildungsniveau der Eltern, abhängen, so dass nicht ohne weiteres klar ist, welche Implikationen mit unterschiedlichen Annahmen über Werte von $\dot{X}$ verbunden sind.[23]

Einige Autoren haben deshalb vorgeschlagen, die Idee einer Intervention definitorisch mit der Vorstellung zu verbinden, dass durch die Intervention exogene Modellvariablen entstehen;[24] so lautet beispielsweise eine

[22] Verbreitet ist die Idee insbesondere bei Autoren, die von Experimenten ausgehen. Systematische Ausarbeitungen der Idee findet man u.a. in Beiträgen von Woodward (2003) und Pearl (2000).

[23] Ohne weiteres sind keine eindeutigen Aussagen möglich, weil innerhalb des Modells $\ddot{Z}$ eine deterministische Variable ist. Könnte man sich stattdessen auf eine stochastische Variable $\dot{Z}$ beziehen, würde sogleich deutlich werden, dass Annahmen über Werte von $\dot{X}$ Implikationen für die Verteilung von $\dot{Z}$ haben.

[24] Woodward (2001: 50); Pearl (1998: 266, 1999: 105-6), Scheines (1997: 189).

Formulierung von Woodward (2001: 50):

„[...] if a variable is endogenous, then intervening on it alters the causal structure of the system in which it figures – giving it a new exogenous causal history.“

Die Idee besteht darin, anhand eines modifizierten Modells ein *hypothetisches Experiment* (Woodward 1999: 201) vorstellbar zu machen. Um den Effekt einer Veränderung $\Delta(x', x'')$ bei einer endogenen Variablen $\dot{X}$ zu untersuchen, wird ein neues Modell gebildet, in dem es die Funktion, die $\dot{X}$ von anderen Modellvariablen abhängig macht, nicht mehr gibt, so dass aus $\dot{X}$ eine exogene Modellvariable $\ddot{X}$ wird.[25] Geht man beispielsweise von dem Modell (7.16b) aus, besteht das hypothetische Experiment darin, stattdessen die Modellvariante (a) zu verwenden, um den Effekt einer Veränderung $\Delta(x', x'')$ zu ermitteln.[26]

2. Ursachen im Kontext von Prozessen. Der eben skizzierte Vorschlag hilft offenbar nicht, wenn man sich dafür interessiert, wie endogenen Variablen im Rahmen eines gegebenen Modells eine kausale Bedeutung gegeben werden kann. Zur Verdeutlichung bleiben wir bei unserem Schulbeispiel. Den Modellvarianten in (7.16) entsprechen unterschiedliche Fragestellungen. Die Modellvariante (a) bezieht sich auf die Frage, wie der schulische Erfolg vom Bildungsniveau der Eltern und vom Schultyp abhängt, und lässt es unbestimmt, wie Werte dieser beiden als „unabhängig“ vorausgesetzten Variablen entstehen. Die Modellvariante (b) bezieht sich dagegen auf einen Prozess, in dem die Wahl eines Schultyps vom Bildungsniveau der Eltern abhängt. Es muss also überlegt werden, wie in diesem Fall von einer kausalen Bedeutung des Schultyps gesprochen werden soll.

Man kann sich auf folgende Frage beziehen: Wie entstehen unterschiedliche Schulabschlüsse? Anhand der Modellvariante (a) kann man antworten, dass der schulische Erfolg vom Bildungsniveau der Eltern *und* vom Schultyp abhängt. Bei der Modellvariante (b) kann man dies zwar auch sagen, muss aber hinzufügen, dass dem Schultyp keine unabhängige, sondern eine vermittelnde Bedeutung zukommt.

Bemerkenswert ist die Veränderung der Fragestellung. Geht man von möglichen Interventionen aus, interessiert man sich für Effekte unterschiedlicher Interventionsmöglichkeiten. Man kann von *praxeologischen Fragestellungen* sprechen, die sich darauf beziehen, wie Menschen durch ihre

[25] Vgl. Woodward (2003: 98), Pearl (2000: 70). Einige unterschiedliche Formulierungsmöglichkeiten werden von Dawid (2002) diskutiert.

[26] Meistens, wie in diesem Beispiel, zeigt sich dann, dass ein solcher Effekt kontextabhängig ist. In der Literatur wird dieses Problem selten explizit diskutiert, weil von abgeschlossenen Modellen ausgegangen wird. Das Modell enthält dann zunächst nur stochastische Variablen, und die Intervention bei einer Modellvariablen $\dot{X}$ führt dazu, dass nur sie durch eine exogene deterministische Variable $\ddot{X}$ ersetzt wird. Im Rahmen eines solchen Modells kann dann einer Veränderung $\Delta(x', x'')$ ein eindeutiger (durchschnittlicher) Effekt zugerechnet werden.

Tätigkeiten bestimmte Wirkungen erzielen können.[27] Die Frage, wie unterschiedliche Schulabschlüsse entstehen, ist offenbar von anderer Art. Sie zielt auf eine kausale Erklärung: man möchte Einsichten in einen Prozess gewinnen, durch den der zu erklärende Sachverhalt entstanden ist oder entstehen könnte.[28] Mögliche Ursachen müssen deshalb in den Prozesskontext eingeordnet werden. Dabei kann es auch relevant sein, auf Akteure Bezug zu nehmen; aber ihre Interventionen bilden dann einen Teil des substantiellen Prozesses, auf den sich die Modellbildung bezieht.

3. Unterschiedliche Arten von Interventionen. Wenn im Hinblick auf funktionale Modelle von Interventionen gesprochen wird, sind zunächst fiktive Interventionen gemeint: Der Modellkonstrukteur trifft willkürlich Annahmen über Werte von Modellvariablen (ggf. im Anschluss an eine vorgängige Veränderung der Modellstruktur). Ob und wie auch von realen Interventionen gesprochen werden kann, hängt dagegen in erster Linie von den Realitätsbezügen des Modells ab. Unsere Überlegungen betreffen Modelle, die sich auf soziale Prozesse beziehen, an denen Akteure beteiligt sind. Es müssen also zwei Arten von Akteuren unterschieden werden:[29]

- *Substantielle Akteure*, die man sich als Träger der substantiellen Prozesse vorstellen kann, durch die die Modellvariablen bestimmte Werte bekommen. Bei unserem Schulbeispiel sind dies zunächst die Schüler und ihre Eltern, durch die Werte von $\dot{X}$ entstehen. An dem Prozess, durch den hinterher Werte von $\dot{Y}$ entstehen, sind natürlich noch andere Akteure beteiligt (Lehrer, Beamte in Schulverwaltungen usw.).
- *Sekundäre Akteure*, auf die man sich bezieht, wenn von Modellkonstruktionen, Beobachtungen, realen und fiktiven Experimenten gesprochen wird.

Zu beachten ist, dass es sich nicht um unterschiedliche Gruppen von Akteuren handelt, vielmehr resultiert die Unterscheidung daraus, dass „man“ sich gedanklich und empirisch auf Prozesse beziehen möchte, an denen Akteure beteiligt sind. Durch die Thematisierung dieser substantiellen Akteure wird „man“ zu einem sekundären Akteur.

Die Unterscheidung zwischen substantiellen und sekundären Akteuren führt nun zu einer Differenzierung im Reden von Interventionen. Einerseits kann man sich auf Interventionen beziehen, die durch die substantiellen Akteure des zu modellierenden Prozessses entstehen; wir sprechen dann von *substantiellen Interventionen*. Andererseits kann man sich auf Inter-

[27] Die Bezeichnung folgt einem Vorschlag von Nurmi (1974); stattdessen könnte auch von poietischen Fragestellungen gesprochen werden.

[28] Kausale Erklärungen setzen nicht unbedingt eine retrospektive Betrachtungsweise voraus, sondern können sich auch auf Ablaufschemas beziehen. Deshalb gibt es hier keine einfache Parallele zu der in Abschnitt 6.1 (§ 5) besprochenen Unterscheidung zwischen retrospektiven und prospektiven Fragestellungen.

[29] Man vgl. hierzu auch Goldthorpe (2001: 8).

ventionen beziehen, die durch sekundäre Akteure zustandekommen; wir sprechen dann von *modalen Interventionen*.[30] So erhält man schließlich auch einen Hinweis, wo modale Interventionen ansetzen können. Bei exogenen Modellvariablen können beliebige Annahmen über ihre Werte bzw. Verteilungen gemacht werden. Für die endogenen Variablen impliziert jedoch das Modell bereits die Existenz von Funktionen, durch die Werte der Variablen enstehen. Werte dieser Variablen können also durch modale Interventionen nicht verändert werden (ohne dem Modell zu widersprechen). Stattdessen müssen modale Interventionen bei den Funktionen ansetzen, durch die Werte der endogenen Variablen entstehen.

4. Modale Interventionen als Verhaltensannahmen. Der mögliche Sinn modaler Interventionen, die bei den Funktionen eines Modells ansetzen, hängt von der inhaltlichen Bedeutung der Funktionen ab. Bei Modellen für soziale Prozesse beziehen sich die Funktionen oft auf das Verhalten (oder Folgen des Verhaltens) der am Prozess beteiligten substantiellen Akteure. Modale Interventionen sind dann gleichbedeutend mit Annahmen über das Verhalten dieser Akteure.

Zur Illustration kann wieder unser Schulbeispiel dienen. Die in diesem Beispiel relevante Verhaltensfunktion betrifft den Zusammenhang zwischen Bildungsniveau der Eltern und Wahl des Schultyps:

$$z \longrightarrow \Pr[\dot{X}|\ddot{Z} = z] \tag{7.17}$$

Eine mögliche Fragestellung bezieht sich auf den durch die Schultypen vermittelten Effekt der unterschiedlichen Bildungsniveaus. Definiert man zur Abkürzung

$$p_{x,0} = \Pr(\dot{X} = 1|\ddot{Z} = 0) \quad \text{und} \quad p_{x,1} = \Pr(\dot{X} = 1|\ddot{Z} = 1)$$

erhält man ausgehend von den Daten in (7.11):

$$\mathrm{E}(\dot{Y}|\ddot{Z} = 1) - \mathrm{E}(\dot{Y}|\ddot{Z} = 0) = 0.3 + 0.1\, p_{x,1} - 0.2\, p_{x,0} \tag{7.18}$$

Also erhöht sich die Differenz in den Wahrscheinlichkeiten für einen erfolgreichen Schulabschluss, wenn Kinder von Eltern mit höherem Bildungsniveau vermehrt den Schultyp 1 besuchen, und sie wird geringer, wenn Kinder von Eltern mit niedrigem Bildungsniveau vermehrt den Schultyp 1 besuchen.

Eine komplementäre Fragestellung betrifft den Effekt der unterschiedlichen Schultypen für den schulischen Erfolg, wenn dabei berücksichtigt wird, dass sich die Zusammensetzung der Schüler in den beiden Schultypen unterscheidet. Ergänzend zu Annahmen über die Verhaltensfunktion

[30] Die Bezeichnung soll daran erinnern, dass diese Interventionen einer Reflexion modaler Fragestellungen dienen.

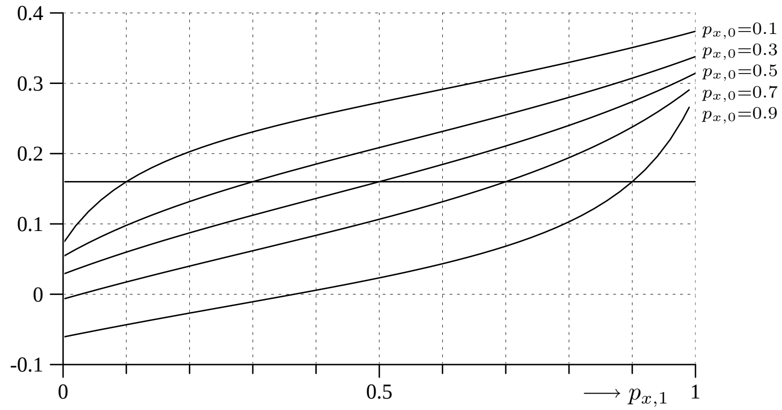


Abb. 7.3-1 Für $p_{x,0} = 0.1, 0.3, 0.5, 0.7$ und 0.9 zeigen die Kurven die Abhängigkeit des Effekts $\mathrm{E}(\dot{Y}|\dot{X} = 1) - \mathrm{E}(\dot{Y}|\dot{X} = 0)$ von den Wahrscheinlichkeiten $p_{x,1}$ (auf der X-Achse).

(7.17) ist dann auch eine Annahme über die Verteilung der Bildungsniveaus der Eltern erforderlich. Nimmt man im Modell (7.16b) anstelle von $\ddot{Z}$ eine Zufallsvariable $\dot{Z}$ an, kann man einen Effekt

$$\mathrm{E}(\dot{Y}|\dot{X} = 1) - \mathrm{E}(\dot{Y}|\dot{X} = 0) \tag{7.19}$$

mithilfe der Formel

$$\mathrm{E}(\dot{Y}|\dot{X} = x) = \frac{\sum_z \mathrm{E}(\dot{Y}|\dot{X} = x, \dot{Z} = z)\Pr(\dot{X} = x|\dot{Z} = z)\Pr(\dot{Z} = z)}{\sum_z \Pr(\dot{X} = x|\dot{Z} = z)\Pr(\dot{Z} = z)}$$

berechnen. Zur Illustration nehmen wir wie in Abschnitt 7.2 (§ 2) an, dass $p_z = \mathrm{P}(Z = 1) = 0.4$ ist. Abbildung 7.3-1 zeigt, wie der durchschnittliche Effekt (7.19) davon abhängt, wie sich die Schulkinder mit unterschiedlichem Bildungsniveau ihrer Eltern auf die Schultypen verteilen.[31]

Zu beachten ist natürlich, dass es sich sowohl bei der Darstellung (7.18) als auch bei der Abbildung 7.3-1 um einen kombinierten Effekt handelt, der aus unterschiedlichen Bildungsniveaus der Eltern, unterschiedlichen Schultypen und unterschiedlichen Verteilungen der Schulkinder auf die Schultypen resultiert. Das zu zeigen ist aber gerade erforderlich, wenn man sich an der Frage orientiert, wie unterschiedliche Schulabschlüsse entstehen.

5. Relevanz der substantiellen Akteure. Autoren, die sich an Interventionen orientieren, unterscheiden oft aktive und passive Prognosen. Zum Beispiel schreibt Woodward (2003: 47):

[31] Zusätzlich eingezeichnet ist der in Abschnitt 7.2 (§ 4) berechnete durchschnittliche Effekt von 0.16, der einer Randomisierung ($p_{x,0} = p_{x,1}$) entsprechen würde.

„It is important to understand that (i) the information that a variable has been set to some value by an intervention is quite different from (ii) the information that the variable has taken that value as the result of some process that leaves intact the causal structure that has previously generated the values of that variable.“

Ein Standardbeispiel ist die Verwendung eines Barometers. Beobachtet man, dass sich der Luftdruck verändert, kann man diese Information für Voraussagen von Wetteränderungen verwenden; aber durch die willlkürliche Veränderung der Anzeige eines Barometers kann man nicht das Wetter verändern. Bei Variablen, deren Werte durch substantielle Akteure entstehen, muss die Überlegung jedoch modifiziert werden. Beobachtungen liefern dann Informationen über die Resultate von substantiellen Interventionen, und Interventionen im Sinne Woodwards müssen explizit als modale Interventionen konzipiert werden.

Das Vorhandensein substantieller Akteure erzeugt nicht nur Sinngrenzen für modale Interventionen, sondern auch für die Randomisierungsidee. Denn diese Idee ersetzt die Entscheidungen der substantiellen Akteure durch Realisationen eines Zufallsgenerators. Die Annahme, dass Randomisierung sinnvoll ist,[32] ist insofern gleichbedeutend mit der Annahme, dass sich das Modell auf einen substantiellen Prozess beziehen soll, an dem keine zu eigenen Entscheidungen fähigen Akteure beteiligt sind.

Schließlich gibt es noch eine weitere methodisch wichtige Konsequenz. Denn man kann bei sozialwissenschaftlichen Anwendungen meistens annehmen, dass die Akteure der substantiellen Prozesse die relevanten Kontextbedingungen besser kennen als die sekundären Akteure, die versuchen, Aspekte der Prozesse durch Modelle zu erfassen.[33] Infolgedessen muss man meistens davon ausgehen, dass es mindestens eine aus der Sicht der sekundären Akteure unbeobachtete Variable gibt, die von den substantiellen Akteuren für Entscheidungen über Interventionen verwendet werden kann, von denen Werte der jeweils beobachteten Variablen abhängen. Es ist jedoch nicht erforderlich, daraus zugleich die Konsequenz zu ziehen, dass man grundsätzlich nur „verzerrte“ Effekte schätzen kann.[34] Vielmehr zeigt sich, dass Ursachen, die durch substantielle Akteure zustandekommen, meistens mit unterschiedlichen Kontexten verbunden sind, so dass abstrakte ceteris-paribus-Überlegungen nicht hilfreich sind.

6. Verzerrungen durch Selbstselektion? Die Vorstellung, dass durch Selbstselektion verzerrte Schätzwerte entstehen können, ist insbesondere in der ökonometrischen Literatur verbreitet.[35] Dabei ist es jedoch wichtig, eine

[32] Oftmals gehört diese Annahme bereits zur Konzeption hypothetischer Experimente; man vgl. z.B. Woodward (2003: 96).

[33] Darauf hat auch Heckman (1996: 461) hingewiesen.

[34] Dies wird beispielsweise von Sobel (2005: 117) nahegelegt.

[35] Man vgl. beispielsweise Maddala (1977, 1983).

Unterscheidung zu beachten.

a) Einerseits kann sich das Selektionsproblem auf einen datenerzeugenden Prozess beziehen. Man kann sich dann an einem Bild der folgenden Art orientieren:

$$Y \xrightarrow{\text{datenerzeugender Prozess}} Y^*$$

In diesem Fall gibt es eine (oft mehrdimensionale) statistische Variable Y, an deren Verteilung $\mathrm{P}[Y]$ man interessiert ist; der datenerzeugende Prozess liefert jedoch nur Daten über eine Variable Y^* mit einer von $\mathrm{P}[Y]$ verschiedenen Verteilung $\mathrm{P}[Y^*]$. Die aus den Daten ermittelbare Verteilung $\mathrm{P}[Y^*]$ liefert dann verzerrte Schätzwerte für die durch $\mathrm{P}[Y]$ definierten Größen, die man schätzen möchte.[36] — In diesem Zusammenhang kann man von Selbstselektion sprechen, wenn der datenerzeugende Prozess von Entscheidungen substantieller Akteure abhängt. Als Beispiel kann man daran denken, dass mit einer Umfrage Daten über Haushaltseinkommen gewonnen werden sollen. Dann ist $\mathrm{P}[Y]$ die für eine bestimmte Gesamtheit von Haushalten definierte Verteilung der Haushaltseinkommen, die man schätzen möchte; die Daten, also Werte von Y^*, hängen jedoch davon ab, ob und wie die Befragungspersonen über ihr Haushaltseinkommen berichten.

b) Eine andere Betrachtungsweise ist erforderlich, wenn man kausale Zusammenhänge untersuchen möchte, an deren Zustandekommen Akteure beteiligt sind. Man kann wieder an das Schulbeispiel denken. Selbstselektion bezieht sich in diesem Fall darauf, dass Werte der Schultyp-Variablen durch Entscheidungen der Schüler bzw. ihrer Eltern entstehen. Dies ist jedoch kein datenerzeugender, sondern ein substantieller Prozess, durch den reale (als Ursachen zu interpretierende) Sachverhalte entstehen.

Auch das Schulbeispiel kann so betrachtet werden, als ob durch Selbstselektion (Entscheidungen der Schüler und Eltern) verzerrte Stichproben entstehen. Geht man von den für eine Gesamtheit Ω definierten statistischen Variablen Z (Bildungsniveau der Eltern), X (Schultyp) und Y (schulischer Erfolg) aus, entstehen durch die Selbstselektion unterschiedliche Verteilungen: $\mathrm{P}[Z|X=0] \neq \mathrm{P}[Z|X=1]$. Verwendete man nur die Daten für die Schüler eines Schultyps, würde eine verzerrte Schätzung der Verteilung der Bildungsniveaus der Eltern für die Gesamtheit Ω resultieren.

Der Frage, wie einer Differenz $\Delta(x', x'')$ beim Schultyp ein bestimmter kausaler Effekt zugerechnet werden kann, enspricht jedoch zunächst noch kein bestimmtes Schätzproblem. Man könnte festlegen, dass der in (7.13) definierte Effekt $\Delta^a(Y|x', x''; Z)$ geschätzt werden soll; dann würde eine

[36] Dem Zweck, solche Verzerrungen reflektierbar zu machen, dienen (probabilistische) Selektionsmodelle; einen Überblick findet man bei Pötter (2006).

einfache Mittelwertdifferenz $\mathrm{M}(Y|X = x'') - \mathrm{M}(Y|X = x')$ nur eine verzerrte Schätzung liefern. *Dieses* Problem entsteht jedoch nur, weil man einen durch $\Delta^a(Y|x', x''; Z)$ definierten Effekt schätzen möchte.

Tatsächlich ist es gar nicht erforderlich, $\Delta^a(Y|x', x''; Z)$ zu schätzen, um die zunächst wesentlichen Einsichten zu gewinnen: dass sowohl die Verteilung auf Schultypen als auch die Effekte unterschiedlicher Schultypen vom Bildungsniveau der Eltern abhängen. Um überlegen zu können, welche Rolle die unterschiedlichen Schultypen für die Herausbildung von Schulabschlüssen spielen, ist eine Kenntnis des durchschnittlichen Effekts $\Delta^a(Y|x', x''; Z)$ gar nicht hilfreich, vielmehr benötigt man Informationen über beide Funktionen, die im Modell (7.16b) den kausalen Prozess charakterisieren.

7. Selektionsprobleme durch Erwartungen? In der ökonometrischen Literatur, die sich mit Fragen der Selbstselektion beschäftigt, wird oft mit Erwartungen argumentiert. Zur Illustration beziehe ich mich auf ein Beispiel von Maddala (1970):

„Returns to college education: If we are given data on incomes of a sample of individuals, some of whom have college education and others not, we have to take account in our analysis the fact that those who have college education are those that chose to go to college and those that do not have college education are those that have chosen (for some reason) not to go to college. This is what we call “self-selectivity”. A naive and commonly used way of analyzing these differences is to define a dummy variable

$$\begin{aligned} D &= 1 \quad \text{if the individual goes to college} \\ &= 0 \quad \text{otherwise} \end{aligned}$$

and estimate an earnings function with D as an extra explanatory variable. This, however, is not a satisfactory solution to the problem.“ (S. 351)

In seiner Begründung bezieht sich Maddala auf eine Situation, in der es keine Kovariablen gibt, so dass (in Maddalas Notation) $\bar{y}_1 - \bar{y}_2$ die durchschnittliche Differenz der Einkommen zwischen Personen mit und ohne College-Ausbildung ist:

„[...] the estimate $\bar{y}_1 - \bar{y}_2$ can be criticized on grounds that it does not take into account the fact that those who went to college did so precisely because they expected their incomes to be higher than otherwise and those who did not go, chose not to go to college precisely because they did not expect their incomes to be higher by doing so. Thus there is a selectivity bias in the estimate $\bar{y}_1 - \bar{y}_2$.“

Wie aber kann daraus, dass Personen sich von Erwartungen leiten lassen, ein Selektionsproblem entstehen?[37] Man kann versuchen, sich zunächst an

[37]Um Maddalas Gedankengang folgen zu können, nehmen wir an, dass seine irrealen Voraussetzungen zutreffen: dass Personen sich beliebig für oder gegen eine College-Ausbildung entscheiden können und dass sie sich bei dieser Entscheidung nur von Erwartungen über das zukünftige Einkommen leiten lassen.

einem Modell der Form

$$\ddot{D} \longrightarrow \dot{Y}$$

zu orientieren, das der Annahme entspricht, dass das Einkommen $\dot{Y}$ von der Ausbildung abhängt ($\ddot{D} = 1$ wenn eine College-Ausbildung vorliegt, andernfalls $\ddot{D} = 0$). Dieses Modell kann jedoch nicht zu einem Selektionsproblem führen, da sich (unter den genannten Annahmen) alle Personen in gleicher Weise entscheiden würden, abhängig davon, ob $\mathrm{E}(\dot{Y}|\ddot{D} = 1)$ größer oder kleiner als $\mathrm{E}(\dot{Y}|\ddot{D} = 0)$ ist. Es muss also noch mindestens eine weitere Variable $\ddot{Z}$ geben, von deren Werten die Entscheidung für oder gegen eine College-Ausbildung abhängig ist. Man kann zwei Möglichkeiten betrachten:

$$\text{(a)} \quad \begin{tikzcd} \ddot{Z} \arrow[d] & \\ \dot{D} \arrow[r] & \dot{Y} \end{tikzcd} \qquad\qquad \text{(b)} \quad \begin{tikzcd} \ddot{Z} \arrow[d] \arrow[dr] & \\ \dot{D} \arrow[r] & \dot{Y} \end{tikzcd} \tag{7.20}$$

In der Modellvariante (a) ist nur die Entscheidungsvariable $\dot{D}$, nicht jedoch das Einkommen von $\ddot{Z}$ abhängig. Dann unterscheiden sich die Personen mit und ohne College-Ausbildung zwar inbezug auf $\ddot{Z}$, dieser Sachverhalt ist jedoch für den Zusammenhang zwischen College-Ausbildung und Einkommen irrelevant. Wir betrachten also die Variante (b). Diese Variante impliziert jedoch, dass $\ddot{Z}$ und $\dot{D}$ interaktive Bedingungen für das Einkommen sind; d.h. es gibt mindestens zwei Werte z' und z'', so dass gilt:

$$\begin{aligned} &\mathrm{E}(\dot{Y}|\dot{D} = 1, \ddot{Z} = z') > \mathrm{E}(\dot{Y}|\dot{D} = 0, \ddot{Z} = z') \\ &\mathrm{E}(\dot{Y}|\dot{D} = 1, \ddot{Z} = z'') < \mathrm{E}(\dot{Y}|\dot{D} = 0, \ddot{Z} = z'') \end{aligned}$$

Denn andernfalls müssten alle Entscheidungen entweder für oder gegen eine College-Ausbildung getroffen werden. Infolgedessen gibt es keinen eindeutig bestimmten Effekt der College-Ausbildung (der unverzerrt geschätzt werden könnte).

8. Fiktive stochastische Antizipationen. Die Überlegungen im vorangegangenen Paragraphen beruhen auf der Voraussetzung, dass Erwartungen in irgendeiner Weise mit Informationen verknüpft sein müssen, die zum Zeitpunkt der Erwartungsbildung verfügbar sind. Für das Beispiel wurde deshalb angenommen, dass es eine Variable $\ddot{Z}$ gibt, die für das spätere Einkommen tatsächlich oder vermeintlich relevant ist und deren Werte den substantiellen Akteuren bekannt sind, so dass sie ihre Erwartungen und sodann Entscheidungen für oder gegen eine College-Ausbildung von diesen Werten abhängig machen können. Stattdessen nimmt Maddala (in dem zitierten Beispiel) an, dass sich Erwartungen unmittelbar auf die Einkommen beziehen können, die mit bzw. ohne College-Ausbildung erzielt werden können.

Dann gelangt man zunächst zu einer Variante des Theorieansatzes von Holland, bei der mit stochastischen individuellen Effekten operiert wird:

$$\dot{D} \longrightarrow\!\!\!\!\!\rightarrow (\dot{Y}_0, \dot{Y}_1) \tag{7.21}$$

Die stochastischen Variablen erfassen das Einkommen, das mit ($\dot{Y}_1$) bzw. ohne ($\dot{Y}_0$) eine College-Ausbildung erzielt werden könnte. Dann wird angenommen, dass über *individuelle* Werte dieser Variablen Erwartungen gebildet werden können und die Entscheidung für oder gegen eine College-Ausbildung von diesen Erwartungen abhängig gemacht wird. Anstelle des Modells (7.21) wird also ein Modell der Form

$$(\dot{Y}_0, \dot{Y}_1) \longrightarrow \dot{D} \longrightarrow\!\!\!\!\!\rightarrow \dot{Y} \tag{7.22}$$

angenommen, wobei die deterministische Funktion, die $(\dot{Y}_0, \dot{Y}_1)$ mit $\dot{D}$ verknüpft, durch die Entscheidungsregel

$$\dot{D} = \begin{cases} 1 \text{ wenn } \dot{Y}_1 \geq \dot{Y}_0 \\ 0 \text{ andernfalls} \end{cases} \tag{7.23}$$

spezifiziert wird. In diesem Modell sind allerdings $\dot{Y}_0$ und $\dot{Y}_1$ zunächst nur latente Variablen, die in keinem bestimmten Zusammenhang mit der Einkommensvariablen $\dot{Y}$ stehen, sondern wie ein geheimes Orakel nur die Entscheidung für oder gegen eine College-Ausbildung bestimmen. Deshalb wird die Annahme hinzugefügt, dass die Erwartungen im Durchschnitt in Erfüllung gehen,[38] also

$$\begin{aligned} \mathrm{E}(\dot{Y}|\dot{D}=1) &= \mathrm{E}(\dot{Y}_1|\dot{D}=1) \\ \mathrm{E}(\dot{Y}|\dot{D}=0) &= \mathrm{E}(\dot{Y}_0|\dot{D}=0) \end{aligned} \tag{7.24}$$

So entsteht schließlich ein geschlossenes Modell,[39] anhand dessen sich ein scheinbares Selektionsproblem

$$\mathrm{E}(\dot{Y}|\dot{D}=1) - \mathrm{E}(\dot{Y}|\dot{D}=0) \neq \mathrm{E}(\dot{Y}_1) - \mathrm{E}(\dot{Y}_0) \tag{7.25}$$

demonstrieren lässt.[40]

Offenbar kann man einwenden, dass es sich nicht um ein kausal interpretierbares Modell handelt, da der stochastischen Funktion $\dot{D} \longrightarrow\!\!\!\!\!\rightarrow \dot{Y}$ kein

[38] Maddala (1977: 354).

[39] Bei der graphischen Modelldarstellung in (7.22) muss zusätzlich ein Doppelpfeil von $(\dot{Y}_0, \dot{Y}_1)$ nach $\dot{Y}$ eingetragen werden.

[40] Es genügt anzunehmen, dass sich die Verteilungen von $\dot{Y}_0$ und $\dot{Y}_1$ überlappen; andernfalls wäre immer $\dot{D} = 0$ oder $\dot{D} = 1$. Natürlich muss auch angenommen werden, dass die Verteilungen von $\dot{Y}_0$ und $\dot{Y}_1$ in irgendeiner Weise gegeben und bekannt sind, da sie durch Beobachtungen von $(\dot{D}, \dot{Y})$ nicht bestimmt werden können.

kausaler Prozess entspricht. Dies entspricht dem Einwand, der in Abschnitt 7.2 (§ 6) diskutiert wurde. Das Modell kann auch verwendet werden, um zu zeigen, wie beliebig das Argumentieren mit Verzerrungen bei (vermeintlich) kausalen Effekten werden kann. Es genügt, sich einen realisierbaren (simulierbaren) Prozess vorzustellen, der dem Modell entspricht:

(1) Zunächst werden mithilfe von Zufallsgeneratoren Realisierungen y_0 und y_1 der Zufallsvariablen $\dot{Y}_0$ und $\dot{Y}_1$ erzeugt.

(2) Dann entscheidet man sich für eine College-Ausbildung und erhält $\dot{Y} = y_1$, wenn $y_1 \geq y_0$ ist; andernfalls entscheidet man sich gegen eine College-Ausbildung und erhält $\dot{Y} = y_0$.

Also kann man den Effekt einer Differenz $\Delta(0,1)$ bei der Ausbildungsvariablen $\dot{D}$ durch

$$\Delta(\dot{Y}; 0, 1) := \mathrm{E}(\dot{Y}|\dot{D}=1) - \mathrm{E}(\dot{Y}|\dot{D}=0) \tag{7.26}$$

definieren. Diese Definition entspricht der linken Seite von (7.25). Welche Bedeutung kommt nun dem Umstand zu, dass sich diese Definition von der Definition

$$\Delta^e(\dot{Y}; 0, 1) := \mathrm{E}(\dot{Y}_1) - \mathrm{E}(\dot{Y}_0) \tag{7.27}$$

unterscheidet? Tatsächlich entspricht diese Definition einem anderen Modell, bei dem anstelle der endogenen Variablen $\dot{D}$ eine exogene Variable $\ddot{D}$ angenommen wird. Es wird dann eine exogene Intervention im Sinne von Woodward oder Pearl vorgenommen.[41] Bezieht mah sich jedoch auf dieses modifiziete Modell, sind beide Effektdefinitionen äquivalent, so dass man auf diese Weise kein Argument für die Definition (7.27) gewinnt.

Zu überlegen ist vielmehr, welche Definition für das Modell (7.22) angemessen ist, in dem $\dot{D}$ eine endogene Variable ist. Da Werte von $\dot{D}$ durch Entscheidungen substantieller Akteure zustandekommen, kann man sich an den Ausführungen über modale Interventionen in § 4 orientieren. Hier geht es dann darum, die kausalen Implikationen unterschiedlicher Entscheidungsregeln zu reflektieren. Dafür eignet sich jedoch nur die Effektdefinition (7.26), denn die Definition (7.27) setzt bereits qua Definition eine bestimmte Entscheidungsregel (Randomisierung) voraus, die dem Verhalten der substantiellen Akteure widerspricht.

9. Invarianz bei Interventionen? Mit der Idee, kausale Effekte als Folgen von Interventionen zu betrachten, ist bei einigen Autoren noch eine weitere Überlegung verbunden: dass die Funktionen eines Modells, das kausal

[41] Stattdessen kann man sich auch vorstellen, dass die Entscheidungsregel (7.23) durch eine Randomisierung ersetzt wird. Innerhalb des modifizierten Modells gibt es dann eine exogene Zufallsvariable $\ddot{D}$, so dass der Pfeil von $(\dot{Y}_0, \dot{Y}_1)$ nach $\ddot{D}$ entfällt.

interpretierbar sein soll, gegenüber Interventionen invariant sein sollten.[42] Allerdings treten zwei Probleme auf. Das erste Problem resultiert aus dem Umstand, dass sich die Invarianzforderung auf Verwendungsmöglichkeiten eines Modells bezieht; sie kann also gar nicht formuliert werden, ohne sich auf einen Verwendungszusammenhang zu beziehen.

Noch wichtiger ist ein zweites Problem, das bei sozialwissenschaftlichen Anwendungen daraus resultiert, dass sich die Modelle auf soziale Prozesse beziehen, an denen Akteure beteiligt sind. Infolgedessen hängen fast immer einige Modellfunktionen davon ab, wie sich die am Prozess beteiligten Akteure verhalten. Insofern können Modelle nicht invariant bzgl. substantieller Interventionen sein. Aber zugleich wird es auch problematisch, die Invarianzforderung auf modale Interventionen zu beziehen; denn solche Interventionen bestehen oftmals gerade darin, Annahmen über Modellfunktionen (die sich auf das Verhalten substantieller Akteure beziehen) zu verändern.

[42] Hausman (1998: chap. 11), Woodward (1999).

Kapitel 8

Modelle für Ereignisse

In diesem Kapitel werden Modelle besprochen, mit denen das Eintreten von Ereignissen – in einzelnen Situationen oder in zeitlichen Folgen von Situationen – erfasst werden kann. Es handelt sich um Varianten funktionaler Modelle (vgl. Abschnitt 5.2), deren Besonderheiten daraus entstehen, dass mithilfe von Ereignisvariablen auf Ereignisse Bezug genommen wird, die in Situationen eintreten können.

Es gibt drei Abschnitte. Im ersten Abschnitt werden der konzeptionelle Ansatz und einfache Modellvarianten, die ohne eine explizite Zeitachse auskommen, dargestellt. Der zweite Abschnitt beschäftigt sich mit Modellen, in denen es explizite Zeitachsen gibt, so dass sich Ereigniswahrscheinlichkeiten und Werte von Kontextvariablen im Verlauf von Situationen verändern können. Im dritten Abschnitt wird besprochen, wie Modelle mit Ereignisvariablen verwendet werden können, um eine dynamische Variante funktionaler Kausalität zu entwickeln, bei der Ursachen als Ereignisse und Wirkungen als Veränderungen von Ereigniswahrscheinlichkeiten betrachtet werden können.

8.1 Situationen und Ereignisse

In diesem Abschnitt wird besprochen, wie Ereignisse durch Variablen erfasst und mithilfe solcher Ereignisvariablen einfache Varianten funktionaler Modelle für Ereignisse, die ohne eine explizite Zeitachse auskommen, gebildet werden können.

1. Ereignisse und Ereignistypen. Wir werden nicht versuchen, einen allgemeinen Ereignisbegriff zu definieren,[1] sondern zunächst nur durch Beispiele andeuten, auf welche Arten von Ereignissen wir uns beziehen wollen: ein Würfel wird geworfen, eine Fensterscheibe zerbricht, ein Kind wird geboren, eine Person wird arbeitslos, zwei Personen heiraten oder lassen sich scheiden. Stets sind in der menschlichen Erfahrungswelt identifizierbare Vorkommnisse gemeint. Vier Aspekte sind wesentlich: An jedem Ereignis ist mindestens ein Objekt beteiligt, bei dem sich etwas verändert, während das Ereignis stattfindet; jedes Ereignis hat eine zeitliche Ausdehnung; oft kann man sagen, dass ein Ereignis früher oder später stattfand als ein anderes Ereignis; Ereignisse können durch Ereignistypen charakterisiert werden.

Wichtig ist insbesondere die Unterscheidung zwischen Ereignissen und Ereignistypen. Ein Ereignis ist ein bestimmtes empirisch identifizierbares Vorkommnis, zum Beispiel das Ereignis, das darin besteht, dass zwei bestimmte Personen heiraten. Ein korrespondierender Ereignistyp wäre ‘heiraten’. Während ein Ereignis nur einmal eintreten kann, kann es viele Ereignisse geben, die sich als Beispiele des gleichen Ereignistyps charakterisieren lassen. Die Charakterisierung eines Ereignisses durch einen Ereignistyp liefert also keine das Ereignis identifizierende Beschreibung. Dass man von Ereignistypen sprechen kann, ist gleichwohl eine wesentliche Voraussetzung, um sich auf *mögliche* Ereignisse beziehen zu können.

2. Ereignisvariablen. Jetzt soll überlegt werden, wie Variablen konzipiert werden können, um Ereignisse und Bedingungen ihres Eintretens zu erfassen. Als Leitfaden soll die Vorstellung dienen, dass man sich auf Situationen beziehen kann, in denen Ereignisse eintreten können (in einer retrospektiven Betrachtung kann dann festgestellt werden, welche Ereignisse tatsächlich eingetreten sind). Um Situationen einer bestimmten Art zu definieren, müssen folgende Angaben gemacht werden.

a) Es muss die Art der Situationen angegeben werden. Zum Beispiel: Es

[1] Auf einige Probleme wurde bereits in Abschnitt 3.2 (§ 4) hingewiesen). Eine gute Orientierung liefert der Beitrag von P. M. S. Hacker (1982).

sollen Situationen betrachtet werden, in denen sich Autofahrer einer Ampel nähern. Oder: Es sollen Situationen betrachtet werden, in denen ein Würfel geworfen wird. Im Folgenden sprechen wir von einer *generischen Situation*, wenn wir uns auf eine (irgendeine) Situation eines bestimmten Typs beziehen.

b) Es muss festgelegt werden, von welcher Art die Ereignisse sind, die in den Situationen eintreten können.[2] Zu diesem Zweck werden *Ereignisvariablen* verwendet, die durch $\dot{E}$ (oder $\dot{E}_1, \dot{E}_2, \ldots$) bezeichnet werden.[3] Der Wertebereich einer Ereignisvariablen wird durch $\tilde{\mathcal{E}}$ (bzw. $\tilde{\mathcal{E}}_1, \tilde{\mathcal{E}}_2, \ldots$) bezeichnet und hat allgemein die Form

$$\tilde{\mathcal{E}} = \{0, 1, \ldots, m\}$$

Die Elemente $1, \ldots, m$ geben die möglichen Ereignistypen an; um sich nur auf diese Elemente zu beziehen, wird die Bezeichnung $\tilde{\mathcal{E}}^* = \tilde{\mathcal{E}} \setminus \{0\}$ verwendet. Das Element 0 bezeichnet keinen Ereignistyp, sondern ist manchmal erforderlich, um auszudrücken, dass (noch) kein Ereignis eingetreten ist.[4] Bezieht man sich auf Situationen, in denen sich Autofahrer einer Ampel nähern, könnte zum Beispiel eine Ereignisvariable $\dot{E}$ mit dem Wertebereich $\tilde{\mathcal{E}} = \{0, 1, 2\}$ und folgenden Bedeutungen definiert werden: $1 \equiv$ ‘der Autofahrer hält an’ und $2 \equiv$ ‘der Autofahrer hält nicht an’.

Es wird angenommen, dass sich die Ereignistypen, die durch den Wertebereich einer Ereignisvariablen unterschieden werden, wechselseitig ausschließen. Für eine Situation können jedoch zwei oder mehr unterschiedliche Ereignisvariablen definiert werden, so dass Ereignisse in unterschiedlichen Kombinationen eintreten können (Beispiele werden in § 5 und im nächsten Abschnitt besprochen). Insbesondere kann anstelle einer Ereignisvariablen $\dot{E}$ mit einem Wertebereich $\{0, 1, \ldots, m\}$ stets auch eine m-dimensionale Ereignisvariable $(\dot{E}_1, \ldots, \dot{E}_m)$ mit einem Wertebereich $\{0, 1\}^m$ verwendet werden, so dass jede Komponente nur einem Ereignistyp entspricht: $(\dot{E}_j = 1) \equiv (\dot{E} = j)$.

c) Es muss angegeben werden, ob und ggf. wie die Situationen durch weitere Merkmale charakterisiert werden können.[5] Variablen, durch die solche Merkmale erfasst werden, nennen wir *Kontextvariablen*. Wir

[2] Eine Formulierung dieser Art setzt offenbar voraus, dass man sich nicht unmittelbar auf reale Situationen, sondern auf ein Ablaufschema (und ein sich daran anschließendes funktionales Modell) beziehen möchte.

[3] In dieser Formulierung handelt es sich um stochastische Modellvariablen. Analog werden die Notationen $\ddot{E}, \ddot{E}_1, \ddot{E}_2, \ldots$ verwendet, um deterministische Ereignisvariablen zu bezeichnen.

[4] Wenn im Folgenden davon gesprochen wird, dass eine Ereignisvariable *einen bestimmten Wert* annimmt, ist deshalb stets ein Wert ungleich Null, der sich auf einen bestimmten Ereignistyp bezieht, gemeint.

[5] Das ist natürlich nur erforderlich, wenn es sich um Merkmale handelt, die bei den

sprechen von *statischen Kontextvariablen*, wenn sich ihre Werte nicht verändern, während die Situation andauert. Dabei kann es sich um Ereignisvariablen handeln, die in einer früheren Situation einen bestimmten Wert angenommen haben, oder um Zustandsvariablen, wie sie für die Formulierung funktionaler Modelle in Kapitel 5 verwendet wurden. Wenn sich Merkmale einer Situation im Verlauf der Situation verändern können, sprechen wir von *dynamischen Kontextvariablen*. Wie Modelle mit solchen Variablen konstruiert werden können, wird in Abschnitt 8.2 besprochen.

Weitere Überlegungen betreffen die zeitliche Ausdehnung von Situationen. Wir legen die Vorstellung zugrunde, dass eine Situation solange andauert, bis zum ersten Mal ein Ereignis eintritt; sobald dies geschieht, entsteht dadurch eine sich anschließende neue Situation.

Ereignisvariablen haben also nicht immer irgendeinen bestimmten (auf ein Ereignis verweisenden) Wert. Dies unterscheidet sie von *Zustandsvariablen*. Zur Illustration kann das Glühlampenbeispiel aus Abschnitt 5.1 (§ 1) dienen. In diesem Beispiel gibt es drei Zustandsvariablen: $\ddot{Y}$ erfasst den Zustand der Glühbirne (sie leuchtet oder nicht), $\ddot{X}$ erfasst den Zustand des Schalters (er ist offen oder geschlossen), und $\ddot{Z}$ erfasst den Zustand der Batterie (sie kann Strom abgeben oder nicht). Alle drei Variablen haben stets einen bestimmten (substantiell interpretierbaren) Wert. Dagegen gehört zu jeder Ereignisvariablen eine Situation, in der die Variable einen bestimmten Wert annehmen kann (z.B. eine Situation, in der ein Schalter geöffnet oder geschlossen werden kann). Bis dies geschieht, hat die Ereignisvariable keinen bestimmten Wert;[6] und hinterher kann sich ihr Wert nicht mehr verändern, da die Tatsache, dass das Ereignis eingetreten ist, nicht mehr verändert werden kann.

3. Daten über Situationen und Ereignisse. Erst wenn Ereignisse eingetreten sind, kann ihr Eintreten festgestellt und können ihre Eigenschaften ermittelt werden. Die dafür vorauszusetzende retrospektive Betrachtungsweise erlaubt es, dass man sich auch mit statistischen Variablen auf Ereignisse (oder im Kontext von Modellen: auf Ereignisvariablen, die bereits bestimmte Werte angenommen haben) beziehen kann. Im einfachsten Fall entspricht jedem Element ω einer Referenzgesamtheit Ω eine bestimmte Situation, und es gibt eine statistische Variable

$$(X, E) : \Omega \longrightarrow \tilde{\mathcal{X}} \times \tilde{\mathcal{E}}$$

wobei sowohl X als auch E mehrdimensional sein können. $X(\omega)$ gibt die

Situationen unterschiedlich sein können. Andernfalls können sie bei der Definition der Art der Situationen, die betrachtet werden sollen, angegeben werden.

[6] Um formale Vollständigkeit zu erreichen, hat die Ereignisvariable bis dahin den Wert Null. Dieser Wert verweist jedoch nicht auf ein Ereignis, sondern nur darauf, dass ein Ereignis stattfinden kann und bisher noch nicht stattgefunden hat.

Eigenschaften der Situation ω an; $E(\omega)$ gibt an, welches Ereignis in der Situation ω eingetreten ist.

4. Funktionale Ereignismodelle. In einer retrospektiven Betrachtungsweise können Ereignisvariablen als (statistische) Variablen konzipiert werden, die bestimmte Werte haben. In einer prospektiven Betrachtungsweise müssen stattdessen Modelle verwendet werden, mit denen über mögliche Ereignisse nachgedacht werden kann. Solche Modelle können als Varianten funktionaler Modelle (wie sie in Kapitel 5 besprochen wurden) konzipiert werden. Wir verwenden allgemein folgende Definition: Ein *Ereignismodell* ist ein funktionales Modell, bei dem mindestens eine endogene Variable eine Ereignisvariable ist.

Von entscheidender Bedeutung ist, das in Ereignismodellen Funktionen mit einer abhängigen Ereignisvariablen stets temporal zu interpretieren sind. Solche Funktionen legen fest, wie in ihnen korrespondierenden Situationen bestimmte Ereignisse (bestimmte Werte der Ereignisvariablen) entstehen können. Infolgedessen sind Modellvariablen, die von einer Ereignisvariablen funktional abhängen, ebenfalls Ereignisvariablen. Als Beispiel kann folgendes Modell dienen:

$$\ddot{E}_0 \xrightarrow{f_1}\!\!\!\!\rightarrow \dot{E}_1 \xrightarrow{f_2}\!\!\!\!\rightarrow \dot{E}_2 \tag{8.1}$$

Bei der Definition der stochastischen Funktion f_1 ist zu beachten, dass $\dot{E}_1$ erst einen bestimmten Wert annehmen kann, nachdem für $\ddot{E}_0$ ein bestimmter Wert entstanden ist. Wir verwenden deshalb die Formulierung

$$j_0 \longrightarrow \Pr[\dot{E}_1 \,\|\, \ddot{E}_0 = j_0]$$

Der Bedingungsstrich $\|$ wird verwendet, um darauf hinzuweisen, dass sich die bedingte Verteilung auf eine Situation bezieht, in der die hinter dem Bedingungsstrich angegebenen Bedingungen bereits realisierte Sachverhalte sind.[7] Analog wird für die stochastische Funktion f_2 die Formulierung

$$j_1 \longrightarrow \Pr[\dot{E}_2 \,\|\, \dot{E}_1 = j_1]$$

verwendet, um deutlich zu machen, dass ein bestimmter Wert von $\dot{E}_1$ entstanden sein muss, bevor $\dot{E}_2$ einen bestimmten Wert annehmen kann.

5. Modelle für einzelne Situationen. Wir besprechen im Folgenden einige Varianten einfacher Ereignismodelle, bei denen die endogenen Ereignisvariablen stochastisch sind. Im einfachsten Fall bezieht sich das Modell nur auf eine Situation, für die zwei Variablen definiert sind: Eine deterministische Kontextvariablen $\ddot{X}$, mit der relevante Merkmale der Situation erfasst

[7] Die Unterscheidung zwischen mit $|$ oder mit $\|$ gebildeten bedingten Verteilungen betrifft also nur die Bedeutung der Bedingungsrelation. In formaler Hinsicht gibt es keinen Unterschied und es können die gleichen Rechenregeln verwendet werden.

werden,[8] und eine stochastische Ereignisvariable $\dot{E}$ für die in der Situation möglichen Ereignisse. Dann wird eine stochastische Funktion

$$x \longrightarrow \Pr[\dot{E} \,\|\, \ddot{X} = x] \tag{8.2}$$

angenommen, die angibt, wie die Wahrscheinlichkeiten für das Eintreten von Ereignissen von Werten der Kontextvariablen $\ddot{X}$ abhängen.

Für eine Situation können auch mehrere Ereignisvariablen definiert werden. Sie müssen dann als Komponenten einer mehrdimensionalen Variablen betrachtet werden. Mit zwei Ereignisvariablen $\dot{E}_1$ und $\dot{E}_2$ könnte ein einfacher Modellansatz durch das Bild

$$\ddot{X} \longrightarrow\!\!\!\!\rightarrow (\dot{E}_1, \dot{E}_2)$$

bzw. die stochastische Funktion

$$x \longrightarrow \Pr[\dot{E}_1, \dot{E}_2 \,\|\, \ddot{X} = x] \tag{8.3}$$

ausgedrückt werden. Wenn angenommen wird, dass sich beide Ereignisvariablen jeweils nur auf einen Ereignistyp beziehen ($\tilde{\mathcal{E}}_1 = \tilde{\mathcal{E}}_2 = \{0, 1\}$), sind folgende Kombinationen möglich:

$\dot{E}_1$	$\dot{E}_2$	
1	0	zuerst tritt das erste Ereignis ein
0	1	zuerst tritt das zweite Ereignis ein
1	1	beide Ereignisse treten gleichzeitig ein
0	0	kein Ereignis tritt ein

Tatsächlich ist der letzte Fall nicht möglich, da die Situation erst abgeschlossen ist, wenn mindestens ein Ereignis eingetreten ist.[9] Das Beispiel zeigt aber, dass bei mehreren Ereignisvariablen mit der Möglichkeit gerechnet werden muss, dass Ereignisse nicht eintreten.

6. Aufeinander folgende Situationen. Bisher wurden Modelle für einzelne Situationen besprochen. Interessante Erweiterungen ergeben sich, wenn man zwei oder mehr Situationen betrachtet, die aufeinander folgen und sich sukzessive bedingen, so dass die Wahrscheinlichkeiten für das Eintreten von Ereignissen auch davon abhängig sein können, was in früheren Situationen geschehen ist.

Zur Illustration verwenden wir das Schulbeispiel aus Abschnitt 7.2. In diesem Beispiel gibt es zwei aufeinander folgende Situationen σ_1 und σ_2.

[8] Wenn ohne Zusatz von Kontextvariablen gesprochen wird, sind stets statische Kontextvariablen gemeint. Als exogene Variable könnte anstelle einer deterministischen Zustandsvariablen $\ddot{X}$ auch eine deterministische Ereignisvariable verwendet werden. Die Modellsituation beginnt dann damit, dass für diese Ereignisvariable ein bestimmter Wert angenommen wird.

[9] Der gemeinsame Wertebereich ist $(\tilde{\mathcal{E}}_1 \times \tilde{\mathcal{E}}_2)^* = \{(1,0), (0,1), (1,1)\}$.

In der Situation σ_1 wird in Abhängigkeit vom Bildungsniveau der Eltern entschieden, welchen Schultyp der Schüler besuchen soll. Es gibt also eine deterministische Kontextvariable $\ddot{Z}$ mit den Werten 0 (niedriges) und 1 (höheres Bildungsniveau) und eine Ereignisvariable $\dot{E}_1$ mit den Werten

$$\dot{E}_1 = \begin{cases} 1 & \equiv \text{ der Schüler kommt in eine Schule des Typs 1} \\ 2 & \equiv \text{ der Schüler kommt in eine Schule des Typs 2} \end{cases}$$

In der Situation σ_2 besucht der Schüler eine Schule des zuvor entschiedenen Typs und es wird eine Ereignisvariable $\dot{E}_2$ mit den Werten

$$\dot{E}_2 = \begin{cases} 1 & \equiv \text{ die Schule wird nicht erfolgreich abgeschlossen} \\ 2 & \equiv \text{ die Schule wird erfolgreich abgeschlossen} \end{cases}$$

betrachtet. Der Kontext wird jetzt durch das Bildungsniveau der Eltern und den Schultyp bestimmt. Wie dieser Kontext durch Variablen erfasst werden kann, ist jedoch davon abhängig, ob man die Situation isoliert oder als eine nachfolgende Situation (also als Teil eines Prozesses) betrachtet. Bei einer isolierten Betrachtung können das Bildungsniveau der Eltern und der Schultyp durch exogene deterministische Variablen erfasst werden. In folgendem Bild kann man sich an der Variante (a) orientieren.[10]

$$\text{(a)}\quad \begin{tikzcd} \ddot{Z} \arrow[dr, twoheadrightarrow] & \\ & \dot{E}_2 \\ \ddot{E}_1 \arrow[ur, twoheadrightarrow] & \end{tikzcd} \qquad \text{(b)}\quad \begin{tikzcd} \ddot{Z} \arrow[dr, twoheadrightarrow] \arrow[dd, twoheadrightarrow] & \\ & \dot{E}_2 \\ \dot{E}_1 \arrow[ur, twoheadrightarrow] & \end{tikzcd} \tag{8.4}$$

In dieser Variante ist $\ddot{E}_1$ eine deterministische exogene Variable, die den Schultyp angibt.[11] Dagegen ist $\dot{E}_1$ in der Variante (b) eine endogene Variable. Das heißt aber auch, dass sich diese Modellvariante auf zwei aufeinander folgende Situationen bezieht. Für die Situation σ_1 gibt es eine stochastische Funktion

$$z \longrightarrow \Pr[\dot{E}_1 \,\|\, \ddot{Z} = z]$$

die zeigt, wie in dieser Situation Werte von $\dot{E}_1$ entstehen können. Für die folgende Situation σ_2 gibt es dann eine stochastische Funktion

$$(z, j_1) \longrightarrow \Pr[\dot{E}_2 \,\|\, \ddot{Z} = z, \dot{E}_1 = j_1]$$

für deren Definition vorausgesetzt wird, dass ein Ereignis $\dot{E}_1 = j_1$ bereits

[10] Die beiden Varianten entsprechen denjenigen in (7.16) in Abschnitt 7.2.

[11] Da $\ddot{E}_1$ eine Ereignisvariable ist, kann man sich auch eine Situation vorstellen, in der $\ddot{E}_1$ einen bestimmten Wert annimmt. Da $\ddot{E}_1$ eine exogene Variable ist, bildet diese Situation aber keinen Teil des Modells und muss deshalb nicht explizit repräsentiert werden.

eingetreten ist. Das ist möglich, denn sobald eine Ereignisvariable in einer Situation einen bestimmten Wert angenommen hat, kann sie in allen folgenden Situationen als eine Kontextvariable verwendet werden.

7. Welche Situationen können entstehen? Bei dem eben betrachteten Beispiel können sich die Situationen, die auf die Ausgangssituation σ_1 folgen, zwar durch Werte ihrer Kontextvariablen unterscheiden; in allen möglichen Folgesituationen gibt es jedoch die gleiche Ereignisvariable $\dot{E}_2$. Wenn es in einer Situation mehrere Ereignisvariablen gibt, müssen auch Verzweigungen in Betracht gezogen werden.

Zur Illustration verwenden wir ein Beispiel, bei dem sich die Situationen auf nichteheliche Lebensgemeinschaften beziehen. Es gibt zwei Ereignisvariablen: Eine Variable $\dot{E}_1$, die den Wert 1 (wenn das Paar heiratet) oder den Wert 2 (wenn das Paar sich trennt oder einer der Beteiligten stirbt) annehmen kann, und eine Variable $\dot{E}_2$, die den Wert 1 annimmt, wenn eine Schwangerschaft eintritt. Außerdem gibt es eine Kontextvariable $\ddot{Z}$, die die Situationen charakterisiert.[12] In diesem Beispiel hängt die Folgesituation auch davon ab, bei welcher der beiden Ereignisvariablen zuerst ein Ereignis stattfindet. Man kann ein Bild der folgenden Art verwenden:

$$\begin{tikzcd} & & \dot{E}_1 \\ \ddot{Z} \arrow[r, "\sigma_1"] & (\dot{E}_1, \dot{E}_2) \arrow[ur, "\sigma_{21}"] \arrow[dr, "\sigma_{22}"'] & \\ & & (\dot{E}_2, \dot{E}_3) \end{tikzcd} \tag{8.5}$$

Für die Situation σ_1 gibt es eine stochastische Funktion

$$z \longrightarrow \Pr[\dot{E}_1, \dot{E}_2 \,\|\, \ddot{Z} = z]$$

die festlegt, wie bei $\dot{E}_1$ oder $\dot{E}_2$ oder auch bei beiden Variablen gleichzeitig bestimmte Werte entstehen können. Wenn zuerst eine Schwangerschaft eintritt, entsteht eine Situation σ_{21}, in der das Paar heiraten oder sich trennen kann. Dieser Situation entspricht die stochastische Funktion

$$z \longrightarrow \Pr[\dot{E}_1 \,\|\, \ddot{Z} = z, \dot{E}_2 = 1]$$

Wenn dagegen zuerst ein Ereignis bei der Variablen $\dot{E}_1$ eintritt, gibt es zwei Möglichkeiten. Wenn das Ereignis $\dot{E}_1 = 2$ (Trennung) eintritt, gibt es in diesem Modell keine Folgesituation. Wenn dagegen das Ereignis $\dot{E}_1 = 1$ (Heirat) eintritt, entsteht eine Situation σ_{22}, in der eine Schwangerschaft und/oder eine Trennung eintreten kann. Um die Möglichkeit einer Trennung zu erfassen, kann jedoch nicht die Variable $\dot{E}_1$ verwendet werden, da sie bereits einen bestimmten Wert ($\dot{E}_1 = 1$) angenommen hat, den sie in einer späteren Situation nicht mehr verändern kann. Will man deshalb diese Möglichkeit explizit in Betracht ziehen, muss eine weitere Ereignisvariable $\dot{E}_3$ definiert werden, durch die erfasst werden kann, ob eine Trennung

[12] Eine empirische Untersuchung solcher Situationen wurde von Blossfeld u.a. (1999) publiziert.

($\dot{E}_3 = 1$) stattfindet.[13] Für die Situation σ_{22} kann dann eine stochastische Funktion

$$z \longrightarrow \Pr[\dot{E}_2, \dot{E}_3 \,\|\, \ddot{Z} = z, \dot{E}_1 = 1]$$

verwendet werden. Natürlich ist es auch möglich, dass in der Situation σ_1 bei beiden Ereignisvariablen ein Ereignis eintritt. Dann entsteht jedoch bei diesem Modellansatz keine explizite Folgesituation.

Dieses Beispiel zeigt, dass in einem Ereignismodell für jede endogene Ereignisvariable und jede Situation, in der sie einen bestimmten Wert annehmen kann, eine separate Funktion angegeben werden muss.[14] Es illustriert auch die Unterschiede bei der Bildung bedingter Verteilungen. Zum Beispiel ist $\Pr(\dot{E}_1 = 1 \,\|\, \ddot{Z} = z, \dot{E}_2 = 1)$ die Wahrscheinlichkeit für eine Heirat in einer Situation, in der eine Schwangerschaft bereits eingetreten ist. Dagegen bezieht sich $\Pr(\dot{E}_1 = 1 | \ddot{Z} = z, \dot{E}_2 = 1)$ auf die Situation σ_1 und gibt die Wahrscheinlichkeit an, dass unter der Bedingung des Eintritts einer Schwangerschaft gleichzeitig geheiratet wird.

8.2 Ereignismodelle mit Zeitachsen

1. Zeitachsen für Situationen. Ereignismodelle, wie sie im vorangegangenen Abschnitt besprochen wurden, beziehen sich auf generische Situationen (Situationen eines bestimmten Typs), die den Kontext bilden, in dem Ereignisse der jeweils thematisierten Art stattfinden können. Es wurde angenommen, dass eine Situation solange besteht, bis zum ersten Mal ein Ereignis eintritt (wodurch dann eine neue Situation entsteht). Das ist ausreichend, wenn man sich nur dafür interessiert, mit welchen Wahrscheinlichkeiten Ereignisse eintreten. Oft interessiert jedoch auch, wie lange es dauert, bis ein Ereignis eintritt, und wie Ereigniswahrscheinlichkeiten von

[13] Man hätte auch von vornherein eine dreidimensionale Variable verwenden können, deren Komponenten jeweils nur einen Ereignistyp erfassen, zum Beispiel: $(\dot{E}_1, \dot{E}_2, \dot{E}_3)$, wobei $\dot{E}_1 = 1$ eine Heirat, $\dot{E}_2 = 1$ eine Schwangerschaft und $\dot{E}_3 = 1$ eine Trennung erfasst. Die in der ersten Situation möglichen Ereignisse wären dann:

$$(1,0,0), (1,1,0), (0,1,0), (0,1,1), (0,0,1)$$

In einigen Fällen könnten sich dann Folgesituationen anschließen, in denen weitere Ereignisse eintreten können.

[14] Wird auf die Kontextvariable $\ddot{Z}$ verzichtet, kann man sich zur Illustration folgende Werte vorstellen:

$$\begin{array}{ll}
\sigma_1 : \Pr(\dot{E}_1 = 1, \dot{E}_2 = 0) = 0.30 & \sigma_{21} : \Pr(\dot{E}_1 = 1 \,\|\, \dot{E}_2 = 1) = 0.6 \\
\sigma_1 : \Pr(\dot{E}_1 = 2, \dot{E}_2 = 0) = 0.48 & \sigma_{21} : \Pr(\dot{E}_1 = 2 \,\|\, \dot{E}_2 = 1) = 0.4 \\
\sigma_1 : \Pr(\dot{E}_1 = 0, \dot{E}_2 = 1) = 0.15 & \sigma_{22} : \Pr(\dot{E}_2 = 1, \dot{E}_3 = 0 \,\|\, \dot{E}_1 = 1) = 0.69 \\
\sigma_1 : \Pr(\dot{E}_1 = 1, \dot{E}_2 = 1) = 0.03 & \sigma_{22} : \Pr(\dot{E}_2 = 0, \dot{E}_3 = 1 \,\|\, \dot{E}_1 = 1) = 0.28 \\
\sigma_1 : \Pr(\dot{E}_1 = 2, \dot{E}_2 = 1) = 0.03 & \sigma_{22} : \Pr(\dot{E}_2 = 1, \dot{E}_3 = 1 \,\|\, \dot{E}_1 = 1) = 0.03
\end{array}$$

Umständen abhängen, die sich während der den jeweiligen Kontext bildenden Situation verändern können. Dann ist es erforderlich, die zeitliche Ausdehnung der Situationen explizit zu erfassen.

Wir verwenden zu diesem Zweck eine deterministische Modellvariable $\ddot{T}$, die angibt, wieviel Zeit seit Beginn einer Situation vergangen ist. Als Wertebereich wird eine diskrete Zeitachse $\mathcal{T}_0 = \{0, 1, 2, \ldots\}$ angenommen, so dass man sich eine Situation auch als eine Folge von Zeitstellen, auf die durch $\ddot{T} = 0, 1, 2, \ldots$ verwiesen wird, vorstellen kann. Der Beginn einer Situation kann in den meisten Fällen durch Verweis auf ein Ereignis festgelegt werden, durch das die Situation entsteht. In diesen Fällen nehmen wir an, dass die Zeitstelle, in der das Ereignis auftritt, zugleich die Zeitstelle $\ddot{T} = 0$ der sich anschließenden Situation ist. Infolgedessen ist es auch möglich, dass zwei aufeinander folgende Ereignisse in der gleichen Zeitstelle stattfinden. Verwendet man zum Beispiel Tage, könnte eine Person am gleichen Tag krank und wieder gesund werden.

2. Zeitabhängige Ereigniswahrscheinlichkeiten. Es wird nun möglich, die Wahrscheinlichkeiten für das Eintreten von Ereignissen auch von der Zeitdauer, die seit dem Beginn einer Situation vergangen ist, abhängig zu machen. Im einfachsten Fall kann man sich auf eine Ereignisvariable $\dot{E}$ mit einem Wertebereich $\tilde{\mathcal{E}} = \{0, 1, \ldots, m\}$ beziehen, die für eine Situation σ definiert ist. Erfasst $\ddot{T}$ die Zeitdauer, die seit Beginn der Situation vergangen ist, können zeitlich bedingte Ereigniswahrscheinlichkeiten

$$r_j(t) := \Pr(\dot{E} = j \,\|\, \ddot{T} = t) \tag{8.6}$$

betrachtet werden. Die Bedingung $\ddot{T} = t$ nimmt Bezug auf die Zeitstelle t und impliziert, dass die Situation mindestens bis zu dieser Zeitstelle andauert, also nicht schon zuvor (in einer Zeitstelle vor t) ein Ereignis eingetreten ist. Wir nennen die Funktion $r_j(t)$ eine *Risikofunktion*. Sie gibt die Wahrscheinlichkeit an, dass in einer Situation, die mindestens bis zur Zeitstelle t andauert, in dieser Zeitstelle ein Ereignis $\dot{E} = j$ eintritt. Offenbar kann auch eine *zusammengefasste Risikofunktion*

$$r(t) := \sum\nolimits_{j \in \tilde{\mathcal{E}}^*} r_j(t) \tag{8.7}$$

definiert werden. Es ist die Wahrscheinlichkeit, dass die Situation mindestens bis zur Zeitstelle t andauert und in dieser Zeitstelle durch irgendein Ereignis beendet wird.

Da $\ddot{T}$ eine deterministische Variable ist, kann sie nur zur Formulierung von Bedingungen verwendet werden. Ausgehend von (8.6) kann aber auch eine stochastische Variable $\dot{T}$ (mit dem gleichen Wertebereich wie $\ddot{T}$) definiert werden, durch die erfasst wird, wie lange die Situation dauert, bis sie durch irgendein Ereignis endet. Ihre Verteilung kann durch eine Survivorfunktion

$$G(t) := \Pr(\dot{T} \geq t) = \prod_{k=0}^{t-1} (1 - r(k))$$

ausgedrückt werden, wobei $G(0) = \Pr(\dot{T} \geq 0) = 1$ ist, und man erhält den Zusammenhang

$$r(t) = \frac{\Pr(\dot{T} = t)}{\Pr(\dot{T} \geq t)} = \frac{\Pr(\dot{T} = t)}{G(t)}$$

Offenbar kann $(\dot{E}, \dot{T})$ auch als eine stochastische Variante der in Abschnitt **??** besprochenen Verweildauervariablen betrachtet werden. Die Risikofunktionen entsprechen dann formal den für die Verweildauervariable definierten (Übergangs-) Ratenfunktionen.

3. Nichteintreten von Ereignissen. Solange die zeitliche Ausdehnung von Situationen unbestimmt bleibt (wie bei den im vorangegangenen Abschnitt besprochenen Ereignismodellen), können für das Nichteintreten von Ereignissen keine Wahrscheinlichkeiten definiert werden. Anders verhält es sich jedoch, wenn man sich Situationen als Folgen von Zeitstellen vorstellen kann. Ist dann $\dot{E}$ eine Ereignisvariable, kann man die Definition

$$\Pr(\dot{E} = 0 \,\|\, \ddot{T} = t) := 1 - r(t) \tag{8.8}$$

verwenden und als Wahrscheinlichkeit interpretieren, dass in der Zeitstelle t keines der Ereignisse, die durch den Wertebereich von $\dot{E}$ möglich sind, eintritt.

4. Aggregation von Zeitstellen. Geht man von Definitionen zeitabhängiger Ereigniswahrscheinlichkeiten aus, gelangt man durch eine Aggregation von Zeitstellen zu dem in Abschnitt 8.1 verfolgten Ansatz, bei dem von der zeitlichen Ausdehnung von Situationen abstrahiert wird. Die Wahrscheinlichkeit, dass in einer Situation ein Ereignis $\dot{E} = j$ eintritt, erhält man durch

$$\Pr(\dot{E} = j) = \sum\nolimits_{t=0}^{\infty} r_j(t) \Pr(\dot{T} \geq t) \tag{8.9}$$

Um die Berechnung zu illustrieren, betrachten wir eine Ereignisvariable $\dot{E}$ mit dem Wertebereich $\tilde{\mathcal{E}} = \{0, 1, 2\}$, die für eine Situation σ definiert ist. Werden beispielsweise konstante Risikofunktionen $r_1(t) = 0.1$ und $r_2(t) = 0.2$ angenommen, entwickelt sich der Prozess bis zur Zeitstelle

$t = 5$ folgendermaßen:

t	$\Pr(\dot{T} \geq t)$	$\Pr(\dot{E} = 1, \dot{T} = t)$	$\Pr(\dot{E} = 2, \dot{T} = t)$	$\Pr(\dot{T} = t)$
0	1.000	0.100	0.200	0.300
1	0.700	0.070	0.140	0.210
2	0.490	0.049	0.098	0.147
3	0.343	0.034	0.069	0.103
4	0.240	0.024	0.048	0.072
5	0.168	0.017	0.034	0.050

Summiert man die zeitstellenspezifischen Wahrscheinlichkeiten mit der Formel (8.9), erhält man $\Pr(\dot{E} = 1) = 1/3$ und $\Pr(\dot{E} = 2) = 2/3$.

Es ist bemerkenswert, dass eine Risikofunktion $r_j(t)$ nicht ausreicht, um die Ereigniswahrscheinlichkeiten $\Pr(\dot{E} = j)$ zu berechnen. Erforderlich ist außerdem die Survivorfunktion $G(t)$ oder, damit äquivalent, die in (8.7) definierte zusammengefasste Risikofunktion $r(t)$. Infolgedessen ist es möglich, dass sich eine Ereigniswahrscheinlichkeit $\Pr(\dot{E} = j)$ verändert, obwohl sich die Risikofunktion für dieses Ereignis, also $r_j(t)$, nicht verändert. Wenn man etwa in dem eben betrachteten Beispiel anstelle von $r_2(t) = 0.2$ die Risikofunktion $r_2(t) = 0.15$ verwendet, erhält man $\Pr(\dot{E} = 1) = 0.4$ und $\Pr(\dot{E} = 2) = 0.6$.

5. Situationen mit mehreren Ereignisvariablen. Wiederum können auch Situationen betrachtet werden, für die mehrere Ereignisvariablen definiert sind. Zur Illustration verwenden wir noch einmal das Beispiel aus Abschnitt 8.1 (§ 7), bei dem sich die Ausgangssituation auf nichteheliche Lebensgemeinschaften bezieht, in denen Heiraten, Trennungen und Schwangerschaften auftreten können. Zeitabhängige Ereigniswahrscheinlichkeiten können in diesem Beispiel durch

$$r_{j_1 j_2}(t) = \Pr(\dot{E}_1 = j_1, \dot{E}_2 = j_2 \,\|\, \ddot{T} = t) \tag{8.10}$$

definiert werden, zu interpretieren als die Wahrscheinlichkeit, dass in der Zeitstelle t der Ausgangssituation (also solange noch kein Ereignis eingetreten ist) das Ereignis $\dot{E}_1 = j_1$ und/oder das Ereignis $\dot{E}_2 = j_2$ eintritt.

Wiederum kann man über die Zeitstellen aggregieren und gelangt dann zu einem von der zeitlichen Ausdehnung der Situationen abstrahierenden Ansatz wie in Abschnitt 8.1 (§ 7). Man kann zum Beispiel die in der folgenden Tabelle angegebenen konstanten Risikofunktionen annehmen

j_1	j_2	$r_{j_1 j_2}(t)$	$\Pr(\dot{E}_1 = j_1, \dot{E}_2 = j_2)$
1	0	0.10	0.30
2	0	0.16	0.48
0	1	0.05	0.15
1	1	0.01	0.03
2	1	0.01	0.03

und erhält dann, wiederum mit der Formel (8.9), die in der rechten Spalte der Tabelle angegebenen aggregierten Ereigniswahrscheinlichkeiten. Sie entsprechen näherungsweise den für die Illustration in Abschnitt 8.1 (§ 7) verwendeten Wahrscheinlichkeiten.

6. Statische und dynamische Kontextvariablen. In Modellen mit zeitabhängigen Ereigniswahrscheinlichkeiten (Risikofunktionen) können sowohl statische als auch dynamische Kontextvariablen verwendet werden. Eine Modellspezifikation erfordert dann eine Definition kontextabhängiger Risikofunktionen, die für eine Ereignisvariable $\dot{E}$ folgende allgemeine Form haben:

$$(t, z, x_t) \longrightarrow \Pr(\dot{E} = j \,\|\, \ddot{T} = t, \ddot{Z} = z, \ddot{X}_t = x_t)$$

In diesem Ausdruck gibt es eine statische Kontextvariable $\ddot{Z}$, deren Wert sich nicht verändert, während die Situation andauert, und eine dynamische Kontextvariable $\ddot{X}_t$, die in jeder Zeitstelle t einen anderen Wert annehmen kann.[15] Offenbar kann auch $\ddot{T}$ als eine dynamische Kontextvariable aufgefasst werden.

Sobald Ereignisvariablen einen bestimmten Werte angenommen haben, können auch sie als Kontextvariablen verwendet werden. Da sie dann ihren Wert nicht mehr verändern können, handelt es sich um statische Kontextvariablen. Man gelangt dann zu bedingten Wahrscheinlichkeiten der folgenden Form:

$$\Pr(\dot{E} = j \,\|\, \ddot{T} = t, \dot{E}' = j', \ldots$$

In diesem Ausdruck ist $\dot{E}'$ eine Kontextvariable, die den Wert j' angenommen hat. Als eine Bedingung für das Eintreten eines Ereignisses $\dot{E} = j$ wird also angenommen, dass ein Ereignis $\dot{E}' = j'$ bereits eingetreten ist, und zwar vor oder auch in der Zeitstelle t. In jedem Fall impliziert unsere Interpretation des Bedingungsstrichs $\|$, dass $\dot{E}' = j'$ als eine bereits realisierte Bedingung für das mögliche Eintreten von $\dot{E} = j$ betrachtet werden kann.

7. Situationsübergreifende Zeitachsen. Bei den bisher betrachteten Modellen beginnt in jeder neuen Situation eine neue Zeitachse. Manchmal erleichtert es die Modellbildung, wenn stattdessen nur eine Zeitachse $\ddot{T}$ verwendet wird, die in der ersten Situation beginnt und fortlaufend alle weiteren Zeitstellen erfasst. Als Beispiel kann folgendes Modell dienen:

$$\ddot{E}_0 \longrightarrow\!\!\!\!\!\rightarrow \dot{E}_1 \longrightarrow\!\!\!\!\!\rightarrow \dot{E}_2$$

[15] Für die zeitliche Datierung einer dynamischen Kontextvariablen $\ddot{X}_t$ gibt es unterschiedliche Möglichkeiten: Man kann sich auf die jeweils aktuelle Zeitstelle t oder auf eine (unmittelbar) vorausgegangene Zeitstelle beziehen. Die Entscheidung sollte von der inhaltlichen Bedeutung und der zeitlichen Ausdehnung der Zeitstellen abhängig gemacht werden.

Nachdem $\ddot{E}_0$ einen bestimmten Wert angenommen hat, beginnt die erste Situation in der Zeitstelle $\ddot{T} = 0$.[16] Für diese Situation können also unmittelbar Risikofunktionen

$$r_{j_1}(t; j_0) := \Pr(\dot{E}_1 = j_1 \,\|\, \ddot{T} = t, \ddot{E}_0 = j_0)$$

definiert werden. Wird nun die Zeitachse bei einem Ereignis $\dot{E}_1 = j_1$ nicht zurückgesetzt, gibt es zwei unterschiedliche Möglichkeiten, um stochastische Funktionen für $\dot{E}_2$ zu definieren. Man kann Wahrscheinlichkeiten der Form

$$\Pr(\dot{E}_2 = j_2 \,\|\, \ddot{T} = t, \dot{E}_1 = j_1) \tag{8.11}$$

verwenden, bei denen die Bedingung ausdrückt, dass sich der Prozess bis zur Zeitstelle t entwickelt hat und in irgendeiner Zeitstelle kleiner oder gleich t das Ereignis $\dot{E}_1 = j_1$ eingetreten ist. Um explizit zu erfassen, wann das Ereignis $\dot{E}_1 = j_1$ eingetreten ist, ist eine ergänzende Notation erforderlich. Wir verwenden einen Operator $\tau(\dot{E})$, der die Zeitstelle angibt, in der eine Ereignisvariable $\dot{E}$ zum ersten Mal einen bestimmten Wert angenommen hat.[17] Dann kann man anstelle von (8.11) auch einen Ausdruck der Form

$$\Pr(\dot{E}_2 = j_2 \,\|\, \ddot{T} = t, \dot{E}_1 = j_1, \tau(\dot{E}_1) = t_1) \tag{8.12}$$

verwenden (die Formulierung setzt natürlich $t \geq t_1$ voraus), und daran anschließend können auch Risikofunktionen für $\dot{E}_2$ definiert werden:

$$r_{j_2}(t; j_1, t_1) := \Pr(\dot{E}_2 = j_2 \,\|\, \ddot{T} = t_1 + t, \dot{E}_1 = j_1, \tau(\dot{E}_1) = t_1)$$

Der in § 1 erläuterten Konvention entsprechend wird angenommen, dass immer dann, wenn ein Ereignis in einer Zeitstelle t eintritt, die sich anschließende neue Situation in der gleichen Zeitstelle beginnt.

8.3 Dynamische Kausalität

1. Ereignisse als dynamische Ursachen. Bei der Diskussion funktionaler Kausalität in Kapitel 6 wurde von einem komparativen Ursachenbegriff ausgegangen, der sich auf eine Veränderung (Differenz) der Werte einer Zustandsvariablen bezieht. In den meisten Fällen entspricht einer solchen Ursache kein Ereignis (weder in der speziellen Bedeutung einer Zustandsveränderung noch im Sinne eines allgemeinen Ereignisbegriffs). In diesem Abschnitt gehen wir stattdessen von der Idee aus, Ursachen von vornherein als Ereignisse aufzufassen. Wir sprechen dann von einer *dynamischen Kausalitätskonzeption*, womit also zunächst nur gemeint sein soll, dass Ursachen als Ereignisse konzipiert werden.

[16] Dies entspricht unserer Konvention, dass eine neue Situation stets in der Zeitstelle beginnt, in der das die Situation einleitende Ereignis stattfindet oder – bei längeren Ereignissen – aufhört.

[17] Der Ausdruck $\tau(\dot{E})$ hat also nur einen bestimmten Wert, wenn die als Argument verwendete Ereignisvariable $\dot{E}$ bereits einen bestimmten Wert angenommen hat.

Zu überlegen ist, wie von Wirkungen eines Ereignisses gesprochen werden kann. Eine oft sinnvolle Vorstellung besteht darin, dass ein Ereignis einen Prozess auslösen oder den Ablauf eines bereits ablaufenden Prozesses modifizieren kann. Die genaue Bedeutung hängt dann aber davon ab, wie die Prozesse konzeptualisiert werden, auf die man sich beziehen möchte (vgl. Kapitel 3). Wir gehen deshalb im Folgenden von einer engeren Vorstellung aus und beziehen uns wiederum auf Ereignisse.

Allerdings ist es meistens nicht möglich, Wirkungen unmittelbar als Ereignisse aufzufassen, denn in den meisten Fällen besteht die Wirkung einer dynamischen Ursache nur darin, dass sich die Wahrscheinlichkeit für das Eintreten eines oder mehrerer anderer Ereignisse verändert. Wir orientieren uns deshalb an folgender Vorstellung: Die *dynamische Wirkung eines Ereignisses* besteht in der dem Ereignis zurechenbaren Veränderung der Wahrscheinlichkeiten für das Eintreten anderer Ereignisse.

So wird auch deutlich, dass ein Modell erforderlich ist, durch das die möglichen Ereignisse fixiert werden, auf die man sich für eine Betrachtung von Wirkungen (= Veränderungen von Eintrittswahrscheinlichkeiten) beziehen möchte. Wir verwenden im Folgenden Ereignismodelle von der Art, wie sie in den vorangegangenen Abschnitten besprochen worden sind; Ereignisse bestehen also darin, dass in einem Modell definierte Ereignisvariablen bestimmte Werte annehmen.[18]

2. Eine Definition dynamischer Kausalität. Eine weitere Überlegung bezieht sich darauf, wie man einem Ereignis eine bestimmte Wirkung (oder mehrere bestimmte Wirkungen) zurechnen kann. Wir verfolgen hier die Idee, dass dafür ein Vergleich erforderlich ist, der die Frage beinhaltet, was geschehen würde (oder wäre), wenn das als Ursache betrachtete Ereignis nicht eintreten würde (oder eingetreten wäre). Zwar kann auf diese Frage im Allgemeinen keine bestimmte Antwort gegeben werden; im Alltag behilft man sich mit (oft fragwürdigen) Normalitätsannahmen. Im Folgenden verwenden wir funktionale Ereignismodelle, die modale (kontrafaktische) Überlegungen ermöglichen.

Um zu einer genauen Definition zu gelangen, beziehen wir uns auf ein Ereignismodell, in dem es eine Ereignisvariable $\dot{E}_1$ (oder eine deterministische Ereignisvariable $\ddot{E}_1$) gibt, so dass man von einem Ereignis $\dot{E}_1 = j_1$ (oder $\ddot{E}_1 = j_1$) sprechen kann. Um diesem Ereignis mögliche Wirkungen

[18] Andere Möglichkeiten, um Vorstellungen über dynamische Kausalität zu konzeptualisieren, entstehen, wenn man von Modellen mit Zustandsvariablen ausgeht und Ereignisse als Zustandsveränderungen auffasst. Man vgl. hierzu den Beitrag von U. Pötter und H.-P. Blossfeld (2001), in dem auch auf eine umfangreiche Literatur zu diesem Ansatz hingewiesen wird.

zuzurechnen, können alle anderen Ereignisse der Form $\dot{E}_2 = j_2$ betrachtet werden, wenn $\dot{E}_2$ ebenfalls zum vorausgesetzten Modell gehört und es einen gerichteten (also auch temporal zu interpretierenden) Pfad gibt, der von $\dot{E}_1$ zu $\dot{E}_2$ führt. Es gibt dann eine bedingte Wahrscheinlichkeit

$$\Pr(\dot{E}_2 = j_2 \,\|\, \dot{E}_1 = j_1, \ddot{Z} = z)$$

wobei sich $\ddot{Z}$ auf alle Kontextvariablen bezieht, deren Werte für die Wahrscheinlichkeit des Eintretens des Ereignisses $\dot{E}_2 = j_2$ relevant sein können, und es kann folgende Sprechweise definiert werden:

□ Die *dynamische Wirkung* des Ereignisses $\dot{E}_1 = j_1$ im Hinblick auf das mögliche Ereignis $\dot{E}_2 = j_2$ im Kontext $\ddot{Z} = z$ ist

$$\Pr(\dot{E}_2 = j_2 \,\|\, \dot{E}_1 = j_1, \ddot{Z} = z) - \Pr(\dot{E}_2 = j_2 \,\|\, \dot{E}_1 = 0, \ddot{Z} = z) \qquad (8.13)$$

Dabei soll angenommen werden, dass der Ausdruck auf der rechten Seite den Wert Null hat, wenn das Ereignis $\dot{E}_2 = j_2$ nur möglich ist, wenn zuvor die Ereignisvariable $\dot{E}_1$ irgendeinen bestimmten Wert angenommen hat. Natürlich kann es vorkommen, dass eine Bezugnahme auf Kontextvariablen nicht erforderlich ist, so dass von einer kontextunabhängigen Wirkung gesprochen werden kann.

Im Folgenden wird diese Definition anhand unterschiedlicher Beispiele verdeutlicht, und es wird auch besprochen, wie im Anschluss an diese Definition von zeitabhängigen Wirkungen gesprochen werden kann.

3. Betrachtung eines Zufallsgenerators. Als ein einfaches Beispiel, bei dem es keine Kontextabhängigkeiten gibt, wird zunächst ein Zufallsgenerator betrachtet, der durch folgendes Bild angedeutet werden kann:

$$\ddot{E}_0 \xrightarrow{\sigma_1} \dot{E}_1 \xrightarrow{\sigma_2} \dot{E}_2$$

Durch eine exogene Ereignisvariable $\ddot{E}_0$ mit dem Wertebereich $\tilde{\mathcal{E}}_0 = \{0, 1\}$ wird erfasst, ob ein Würfel geworfen wird. Wenn dies geschieht, entsteht eine Situation σ_1, in der die stochastische Ereignisvariable $\dot{E}_1$ mit dem Wertebereich $\tilde{\mathcal{E}}_1 = \{0, 1, \ldots, 6\}$ die resultierende Augenzahl erfasst. Es wird ein normaler Würfel vorausgesetzt, also

$$\Pr(\dot{E}_1 = j \,\|\, \ddot{E}_0 = 1) = 1/6 \quad \text{für } j = 1, \ldots, 6 \qquad (8.14)$$

Schließlich entsteht, wenn $\dot{E}_1$ einen bestimmten Wert angenommen hat, eine weitere Situation σ_2, in der durch die stochastische Ereignisvariable $\dot{E}_2$ mit dem Wertebereich $\tilde{\mathcal{E}}_2 = \{0, 1, 2\}$ erfasst wird, ob in σ_1 eine ungerade ($\dot{E}_2 = 1$) oder gerade Augenzahl ($\dot{E}_2 = 2$) entstanden ist.

Welche Wirkungen können den in diesem Modell möglichen Ereignissen zugerechnet werden? Beginnen wir damit, dass der Würfel geworfen wird: $\ddot{E}_0 = 1$. Das Modell gibt an, mit welchen Wahrscheinlichkeiten dann die Ereignisvariablen $\dot{E}_1$ und $\dot{E}_2$ bestimmte Werte annehmen. Die in § 2 vorgeschlagene Definition erfordert es, dass auch eine Aussage darüber gemacht wird, was geschehen würde, wenn der Würfel nicht geworfen wird. Wie gesagt, kann auf eine Frage dieser Art im Allgemeinen keine bestimmte Antwort gegeben werden. Bezieht man sich jedoch auf das Modell, kann man sagen, dass ohne das Ereignis $\ddot{E}_0 = 1$ gar kein weiteres Ereignis eintreten könnte, also $\Pr(\dot{E}_1 = j \,\|\, \ddot{E}_0 = 0) = 0$ ist. Man kommt zu dem Ergebnis, dass die dynamische Wirkung des Ereignisses $\ddot{E}_0 = 1$ durch die Wahrscheinlichkeiten (8.14) angegeben werden kann.

Betrachten wir jetzt ein endogenes Ereignis, zum Beispiel $\dot{E}_1 = 1$. Das Modell liefert zunächst die dadurch bedingten Wahrscheinlichkeiten

$$\Pr(\dot{E}_2 = 1 \,\|\, \dot{E}_1 = 1) = 1 \quad \text{und} \quad \Pr(\dot{E}_2 = 2 \,\|\, \dot{E}_1 = 1) = 0$$

Sie drücken aber bereits die dynamischen Wirkungen aus, da wiederum $\dot{E}_1 \neq 0$ eine Voraussetzung dafür ist, dass $\dot{E}_2$ einen bestimmten Wert annehmen kann, so dass $\Pr(\dot{E}_2 = 1 \,\|\, \dot{E}_1 = 0) = \Pr(\dot{E}_2 = 2 \,\|\, \dot{E}_1 = 0) = 0$ ist.

Das Beispiel zeigt auch, dass dynamische Wirkungen nicht immer mit positiven Wahrscheinlichkeiten verbunden sein müssen. Denn eine der dem Ereignis $\dot{E}_1 = 1$ zurechenbaren Wirkungen besteht in diesem Beispiel darin, dass die Wahrscheinlichkeit für das Ereignis $\dot{E}_2 = 2$ Null wird.

4. Unterschiedliche modale Vergleiche. Folgt man der Definition in § 2, ist zur Ermittlung der dynamischen Wirkungen eines Ereignisses $\dot{E} = j$ ein Vergleich erforderlich, bei dem eine Situation, in der das Ereignis $\dot{E} = j$ eintritt, mit einer Situation verglichen wird, in der $\dot{E}$ noch keinen bestimmten Wert angenommen hat. Stattdessen kann man auch eine Situation, in der das Ereignis $\dot{E} = j$ eintritt, mit einer Situation vergleichen, in der ein anderes Ereignis $\dot{E} = j'$ eintritt. Aber das wäre dann ein wesentlich anderer Vergleich, und das Ergebnis würde dann auch davon abhängen, mit welchem Alternativereignis der Vergleich durchgeführt wird.

Um die unterschiedlichen Arten der Vergleiche zu verdeutlichen, eignet sich auch das Schulbeispiel aus Abschnitt 8.1 (§ 6). Als Ursachen kann man die Ereignisse $\dot{E}_1 = 1$ (der Schüler kommt in eine Schule des Typs 1) und $\dot{E}_1 = 2$ (der Schüler kommt in eine Schule des Typs 2) betrachten. Folgt man der Definition aus § 2, können jeder dieser Ursachen gesondert dynamische Wirkungen inbezug auf Ereignisse $\dot{E}_2 = j_2$ (erfolgreicher oder nicht erfolgreicher Schulabschluss) zugerechnet werden; und da die Ereignisvariable $\dot{E}_1$ irgendeinen bestimmten Wert annehmen muss, bevor irgendein Ereignis $\dot{E}_2 = j_2$ eintreten kann, können diese Wirkungen einfach durch die kontextabhängigen bedingten Wahrscheinlichkeiten

$$\Pr(\dot{E}_2 = j_2 \,\|\, \dot{E}_1 = 1, \ddot{Z} = z) \quad \text{und} \quad \Pr(\dot{E}_2 = j_2 \,\|\, \dot{E}_1 = 2, \ddot{Z} = z)$$

angegeben werden. Daran anschließend können diese kontextabhängigen Wirkungen der Ereignisse $\dot{E}_1 = 1$ und $\dot{E}_1 = 2$ verglichen werden. Aber ein solcher Vergleich ist für die Feststellung der dynamischen Wirkungen nicht erforderlich; vielmehr müssen umgekehrt zuerst die jeweiligen Wirkungen ermittelt werden, bevor sie verglichen werden können.

Es ist bemerkenswert, dass der Vergleich, auf dem die Kausalitätsdefinition in § 2 beruht, nur möglich ist, wenn man sich auf Ereignisvariablen bezieht. Er kann nicht vorgenommen werden, wenn man stattdessen von Zustandsvariablen ausgeht, da Zustandsvariablen immer irgendeinen bestimmten Wert haben, so dass nur Implikationen unterschiedlicher Werte verglichen werden können (wie bei den Definitionen der Effekte komparativer Ursachen in Kapitel 6).

5. Exogene intervenierende Ursachen. Bisher wurden Beispiele betrachtet, bei denen eine Ereignisvariable $\dot{E}_1$, die zur Definition einer Ursache $\dot{E}_1 = j_1$ verwendet wird, irgendeinen bestimmten Wert annehmen muss, damit ein Ereignis $\dot{E}_2 = j_2$, für das eine Wirkung ermittelt werden soll, eintreten kann. Wenn dies nicht der Fall ist, sprechen wir von *intervenierenden Ursachen.* Folgendes Modell liefert ein einfaches Beispiel:

$$\begin{tikzcd} \ddot{E}_0 \arrow[r, two heads] & \dot{E}_2 \\ \ddot{E}_1 \arrow[ur, two heads] & \end{tikzcd} \tag{8.15}$$

Das Ereignis $\ddot{E}_0 = 1$ soll bedeuten, dass eine Person krank wird. Daran anschließend können zwei andere Ereignisse eintreten: die Person wird gesund ($\dot{E}_2 = 1$) oder sie stirbt ($\dot{E}_2 = 2$). Außerdem kann ein Ereignis $\ddot{E}_1 = 1$ eintreten, das darin besteht, dass eine bestimmte medizinische Behandlung erfolgt. Der Unterschied zu den bisher betrachteten Beispielen besteht darin, dass durch $\dot{E}_2$ definierbare Ereignisse eintreten können, ohne dass zuvor die Ereignisvariable $\ddot{E}_1$ einen bestimmten Wert annehmen muss. Um das Modell zu spezifizieren, müssen also zwei Wahrscheinlichkeitsverteilungen angegeben werden:

$$\Pr(\dot{E}_2 = j_2 \,\|\, \ddot{E}_0 = 1, \ddot{E}_1 = 1) \quad \text{und} \quad \Pr(\dot{E}_2 = j_2 \,\|\, \ddot{E}_0 = 1, \ddot{E}_1 = 0)$$

Aus ihrer Differenz erhält man dann die Wirkungen des Ereignisses $\ddot{E}_1 = 1$ im Hinblick auf die möglichen Ereignisse $\dot{E}_2 = 1$ und $\dot{E}_2 = 2$.

6. Endogene intervenierende Ursachen. In dem Modell (8.15) ist $\ddot{E}_1$ eine exogene Ereignisvariable, so dass Annahmen darüber, ob ein Ereignis $\ddot{E}_1 = 1$ eintritt oder nicht eintritt, beliebig vorgenommen werden können. Stattdessen kann man auch Modelle betrachten, bei denen intervenierende Ursachen durch endogene Ereignisvariablen entstehen können. Zur Illustration verwenden wir noch einmal das Beispiel aus Abschnitt 8.1 (§ 7), bei dem sich die Ausgangssituation auf nichteheliche Lebensgemeinschaften bezieht, in denen Heiraten, Trennungen und Schwangerschaften auftreten können.

Um an die Notationen des vorangegangenen Paragraphen anzuschließen, soll $\ddot{E}_0 = 1$ den Beginn einer nichtehelichen Lebensgemeinschaft erfassen. Durch $\dot{E}_2$ wird erfasst, ob es zu einer Heirat ($\dot{E}_2 = 1$) oder Trennung ($\dot{E}_2 = 2$) kommt, und $\dot{E}_1 = 1$ erfasst, ob eine Schwangerschaft eintritt. Die Fragestellung bezieht sich darauf, wie eine Schwangerschaft die Wahrscheinlichkeit einer Heirat oder einer Trennung verändert. Es handelt sich um eine intervenierende Ursache, da eine Heirat oder eine Trennung auch ohne eine Schwangerschaft eintreten kann.

Wichtig ist jetzt aber, dass $\dot{E}_1$ als eine endogene Ereignisvariable betrachtet werden soll; d.h. wenn eine nichteheliche Lebensgemeinschaft begonnen wurde, kann auch mit einer durch das Modell bestimmten Wahrscheinlichkeit eine Schwangerschaft eintreten. Um diese Möglichkeit anzudeuten, könnte man in der graphischen Darstellung in (8.15) $\dot{E}_1$ anstelle von $\ddot{E}_1$ verwenden und einen zusätzlichen Pfeil von $\ddot{E}_0$ nach $\dot{E}_1$ einzeichnen. Das resultierende Bild wäre jedoch irreführend, weil es nicht deutlich macht, dass in der durch $\ddot{E}_0 = 1$ entstehenden Ausgangssituation sowohl $\dot{E}_1$ als auch $\dot{E}_2$ bestimmte Werte annehmen können. Besser ist deshalb eine graphische Darstellung in der Art von (8.5) in Abschnitt 8.1. Für unsere gegenwärtige Fragestellung kann man die Darstellung

$$\ddot{E}_0 \xrightarrow{\sigma_1} (\dot{E}_1, \dot{E}_2) \xrightarrow{\sigma_2} \dot{E}_2 \tag{8.16}$$

verwenden. In der Situation σ_1 gibt es die Wahrscheinlichkeiten

$$\Pr(\dot{E}_1 = j_1, \dot{E}_2 = j_2 \,\|\, \ddot{E}_0 = 1)$$

Daraus erhält man die Wahrscheinlichkeit für eine Heirat oder für eine Trennung, bevor eine Schwangerschaft eingetreten ist (oder gleichzeitig mit einer Schwangerschaft), durch

$$\begin{aligned} &\Pr(\dot{E}_2 = j_2 \,\|\, \ddot{E}_0 = 1) = \\ &\qquad \Pr(\dot{E}_1 = 0, \dot{E}_2 = j_2 \,\|\, \ddot{E}_0 = 1) + \Pr(\dot{E}_1 = 1, \dot{E}_2 = j_2 \,\|\, \ddot{E}_0 = 1) \end{aligned}$$

Eine für unsere Fragestellung relevante Folgesituation σ_2 entsteht, wenn in der ersten Situation eine Schwangerschaft, aber noch keine Heirat oder Trennung eingetreten ist. Dieser Situation entsprechen die Wahrscheinlichkeiten

$$\Pr(\dot{E}_2 = j_2 \,\|\, \ddot{E}_0 = 1, \dot{E}_1 = 1)$$

Wirkungen einer Schwangerschaft können also durch die Differenz

$$\Pr(\dot{E}_2 = j_2 \,\|\, \ddot{E}_0 = 1, \dot{E}_1 = 1) - \Pr(\dot{E}_2 = j_2 \,\|\, \ddot{E}_0 = 1)$$

ermittelt werden. Verwendet man die zur Illustration in Abschnitt 8.1 (§ 7) angegebenen Zahlen, findet man, dass sich durch eine Schwangerschaft die Wahrscheinlichkeit einer Heirat um $0.6 - 0.33 = 0.27$ erhöht, die Wahrscheinlichkeit einer Trennung um $0.4 - 0.51 = -0.11$ verringert.

7. Zeitabhängige dynamische Wirkungen. Wenn man Ereignismodelle mit Zeitachsen verwendet, kann man auch zeitabhängige dynamische Wirkungen betrachten. Im einfachsten Fall kann man sich auf zwei Ereignisvariablen $\dot{E}_1$ und $\dot{E}_2$ beziehen, und $\dot{E}_2$ kann nur dann einen bestimmten Wert annehmen, wenn zuvor $\dot{E}_1$ irgendeinen Wert angenommen hat. Infolgedessen ist $\Pr(\dot{E}_2 = j_2 \,\|\, \dot{E}_1 = 0) = 0$, und die zeitabhängigen dynamischen Wirkungen eines Ereignisses $\dot{E}_1 = j_1$ können unmittelbar durch eine Risikofunktion

$$\Pr(\dot{E}_2 = j_2 \,\|\, \dot{E}_1 = j_1, \ddot{T} = t)$$

ausgedrückt werden, wobei angenommen wird, dass die Zeitachse $\ddot{T}$ mit dem Eintreten des Ereignisses $\dot{E}_1 = j_1$ beginnt.

Wiederum muss bei intervenierenden Ursachen etwas anders vorgegangen werden. Wir beziehen uns zunächst auf den in § 5 besprochenen Modellansatz. Wenn die exogene Ursache $\ddot{E}_1 = 1$ in einer Zeitstelle $\tau(\ddot{E}_1) = t_1$ auftritt, kann man zunächst eine in der Zeitstelle t_1 beginnende Risikofunktion

$$r_{j_2}(t; 1, t_1) := \Pr(\dot{E}_2 = j_2 \,\|\, \ddot{E}_0 = 1, \ddot{E}_1 = 1, \tau(\ddot{E}_1) = t_1, \ddot{T} = t_1 + t) \tag{8.17}$$

betrachten. Ein sinnvoller Vergleich wird möglich, wenn man annimmt, dass ein Ereignis $\ddot{E}_1 = 1$ bis zur Zeitstelle t_1 einschließlich nicht eingetreten ist, also durch eine Betrachtung der konditionalen Risikofunktion

$$r_{j_2}(t; 0, t_1) := \Pr(\dot{E}_2 = j_2 \,\|\, \ddot{E}_0 = 1, \ddot{E}_1 = 0, \ddot{T} = t_1 + t) \tag{8.18}$$

Die zeitabhängige Wirkung eines Ereignisses $\ddot{E}_1 = 1$, das in der Zeitstelle t_1 auftritt, kann dann durch die Differenz der in (8.17) und (8.18) angegebenen Risikofunktionen definiert werden:

$$\text{Zeitabhängige Wirkung} = r_{j_2}(t; 1, t_1) - r_{j_2}(t; 0, t_1) \tag{8.19}$$

Die Wirkung kann also nicht nur davon abhängig sein, wieviel Zeit seit dem Auftreten der Ursache vergangen ist, sondern auch von t_1, d.h. von der zeitlichen Differenz zwischen dem Ereignis $\ddot{E}_0 = 1$ und dem Auftreten der Ursache $\ddot{E}_1 = 1$.

Ganz analog kann man zeitabhängige Wirkungen endogener intervenierender Ursachen definieren, denn beide Risikofunktionen können auch formuliert werden, wenn anstelle von $\ddot{E}_1$ eine endogene Ereignisvariable $\dot{E}_1$ verwendet wird.

8. Zeitlich lokale und integrierte Wirkungen. Die Definition (8.19) bezieht sich auf Risikofunktionen und liefert zunächst zeitliche lokale Wirkungen. Stattdessen kann man auch fragen, wie eine intervenierende Ursache die

über alle späteren Zeitstellen integrierte Wahrscheinlichkeit für das Eintreten eines anderen Ereignisses verändert. Die Überlegung erfolgt analog zu derjenigen in Abschnitt 8.2 (§ 4).

Zur Illustration betrachten wir nochmal das Modell (8.15). Um die integrierte Wirkung einer in der Zeitstelle t_1 auftretenden Ursache $\ddot{E}_1 = 1$ zu berechnen, benötigt man zusätzlich zu den in (8.17) und (8.18) definierten Risikofunktionen Wahrscheinlichkeiten dafür, dass die in der Zeitstelle t_1 beginnende Situation aufrechterhalten bleibt. Da diese Wahrscheinlichkeiten auch davon abhängen, ob zuvor die Ursache eingetreten ist, sind zwei Verweildauervariablen zu unterscheiden. Eine Verweildauervariable $\dot{T}_1$ mit der Verteilung

$$\Pr(\dot{T}_1 \geq t_1) = \prod_{k=0}^{t-1} (1 - r(t; 1, t_1)$$

wobei $r(t; 1, t_1) := \sum_j r_j(t; 1, t_1)$ ist, erfasst, wielange die Situation aufrechterhalten bleibt, wenn das Ereignis $\ddot{E}_1 = 1$ in der Zeitstelle t_1 eintritt. Analog wird eine Verweildauervariable $\dot{T}_0$ unter der Bedingung gebildet, dass die Ursache nicht eintritt. Die integrierte Wirkung erhält man dann durch

$$\sum\nolimits_{t=0}^{\infty} r_{j_2}(t; 1, t_1) \Pr(\dot{T}_1 \geq t) - r_{j_2}(t; 0, t_1) \Pr(\dot{T}_0 \geq t) \tag{8.20}$$

Zur Illustration knüpfen wir an das Beispiel in Abschnitt 8.2 (§ 4) an, bei dem konstante Risikofunktionen verwendet werden: $r_1(t; 0, t_1) = 0.1$ und $r_2(t; 0, t_1) = 0.2$. Infolgedessen hängt die integrierte Wirkung nicht davon ab, wann die Ursache eintritt, und man kann beispielsweise annehmen, dass sich die Risikofunktionen durch das Auftreten der Ursache folgendermaßen verändern: $r_1(t; 1, t_1) = 0.15$ und $r_2(t; 0, t_1) = 0.25$. Als integrierte Wirkungen der Ursache $\ddot{E}_1 = 1$ erhält man

$$\text{für } \dot{E}_2 = 1\text{:} \quad 0.375 - 1/3 \approx 0.042$$
$$\text{für } \dot{E}_2 = 2\text{:} \quad 0.625 - 2/3 \approx -0.042$$

Das Beispiel zeigt auch erneut, dass es keine einfachen Zusammenhänge zwischen Risikofunktionen und zeitlich integrierten Ereigniswahrscheinlichkeiten gibt.

Literatur

Atteslander, P. 1976. Sozialwissenschaftliche Aspekte von Raumordnung und Raumplanung. In: Ders. (Hg.), Soziologie und Raumplanung, 10–71. Berlin: de Gruyter.

Balzer, W. 1997. Die Wissenschaft und ihre Methoden. Grundsätze der Wissenschaftstheorie. München: Alber.

Bates, F. L., Peacock, W, G. 1989. Conceptualizing Social Structure: The Misuse of Classification in Structural Modeling. American Sociological Review 54, 565–577.

Black, M. 1962. Models and Metaphors. Studies in Language and Philosophy. Ithaca: Cornell University Press.

Blake, J., Davis, K. 1964. Norms, Values, and Sanctions. In: R. E. L. Faris (ed.), Handbook of Modern Sociology, 456–484. Chicago: Rand McNally.

Blau, P. M. 1974. Parameters of Social Structure. American Sociological Review 39, 615–635.

Blau, P. M. 1977. A Macrosociological Theory of Social Structure. American Journal of Sociology 83, 26–54.

Blau, P. M. 1994a. Structural Contexts of Opportunities. Chicago: University of Chicago Press.

Blaut, J. M. 1971. Space, Structure and Maps. Tijdschrift voor Economische en Sociale Geografie 62, 18–21.

Blossfeld, H.-P., Klijzing, E., Pohl, K., Rohwer, G. 1999. Why Do Cohabiting Couples Marry? An Example of a Causal Event History Approach to Interdependent Systems. Quality & Quantity 33, 229–242.

Blossfeld, H.-P., Rohwer, G. 2002. Techniques of Event History Modeling (2nd ed.). Mahwah: Lawrence Erlbaum.

Bodzenta, E. 1979. Sozialstruktur im Wandel – am Beispiel Österreichs. In: K. Salamun (Hg.), Sozialphilosophie als Aufklärung, 379–398. Tübingen: Mohr.

Bossel, H. 1992. Modellbildung und Simulation. Brauschweig: Vieweg.

Boudon, R., Bourricaud, F. 1992. Soziologische Stichworte. Opladen: Westdeutscher Verlag.

Büschges, G., Abraham, M., Funk, W. 1998. Grundzüge der Soziologie. 3. Aufl. München: Oldenbourg.

Campbell, C. 1996a. On the Concept of Motive in Sociology. Sociology 30, 101–114.

Campbell, C. 1996b. The Myth of Social Action. Cambridge: Cambridge University Press.

Cancian, F. M. 1975. What are Norms? A Study of Beliefs and Action in a Maya Community. Cambridge: Cambridge University Press.

Cantor, G. 1962. Gesammelte Abhandlungen mathematischen und philosophischen Inhalts. Hrsg. von E. Zermelo. Hildesheim: Georg Olms.

Carroll, J. W. 1991. Property-Level Causation? Philosophical Studies 63, 245-270.

Cartwright, N. 1979. Causal Laws and Effective Strategies. Noûs 13, 419–437.

Cartwright, N. 2001. Modularity: It Can–and Generally Does–Fail. In: M. C. Galavotti, P. Suppes, D. Costantini (eds.), Stochastic Causality, 65–84. Stanford: Center for the Study of Language and Information.

Clogg, C. C., Haritou, A. 1997. The Regression Method of Causal Inference and a Dilemma Confronting This Method. In: V. R. McKim, S. P. Turner (eds.), Causality in Crisis? Statistical Methods and the Search for Causal Knowledge in the Social Sciences, 83–112. Notre Dame: University of Notre Dame Press.

Collier, D., Brady, H. E., Seawright, J 2004. Sources of Leverage in Causal Inference: Toward an Alternative View of Methodology. In: H. E. Brady, D. Collier (eds.), Rethinking Social Inquiry, 229–266. New York: Rowman & Littlefield.

Cooper, G. F. 1999. An Overview of the Representation and Discovery of Causal Relationships Using Bayesian Networks. In: C. Glymour, G. F. Cooper (eds.), Computation, Causation, and Discovery, 3-62. Cambridge: MIT Press.

Cox, D. R. 1992. Causality: Some Statistical Aspects. Journal of the Royal Statistical Society A 155, 291–301.

Cox, D. R., Wermuth, N. 1996. Multivariate Dependencies – Models, Analysis and Interpretation. London: Chapman & Hall.

d'Alembert, J. 1756. Artikel *Experimentell*. In: Artikel aus der von Diderot und d'Alembert herausgegebenen Enzyklopädie, 433–437. Hrsg. von M. Naumann. Frankfurt: Röderberg 1984.

Danto, A. C. 1965. Analytical Philosophy of History. Dt.: Analytische Philosophie der Geschichte. Frankfurt: Suhrkamp 1980

Daston, L. 1998. Die Kultur der wissenschaftlichen Objektivität. In: O. G. Oexle (Hg.), Naturwissenschaft, Geisteswissenschaft, Kulturwissenschaft: Einheit – Gegensatz – Komplementarität? 9–39. Göttingen: Wallstein.

Dawid, A. P. 2000. Causal Inference without Counterfactuals. Journal of the American Statistical Association 95, 407–424.

Dawid, A. P. 2002. Influence Diagrams for Causal Modelling and Inference. International Statistical Review 70, 161–189.

Dawid, A. P. 2006. Counterfactuals, Hypotheticals and Potential Responses: A Philosophical Examination of Statistical Causality. Research Report No. 269, Department of Statistical Science, University College London.

Edinger, K.-E. 1998. Sozialstruktur, soziale Ungleichheit, soziale Schichtung. Paderborn: Schöningh.

Eells, E. 1988. Probabilistic Causal Interaction and Disjunctive Causal Factors. In: J. H. Fetzer (ed.), Probability and Causality, 189–209. Dordrecht: Reidel.

Eells, E. 1991. Probabilistic Causality. Cambridge: Cambridge University Press.

Eells, E., Sober, E. 1983. Probabilistic Causality and the Question of Transitivity. Philosophy of Science 50, 35–57.

Elwert, F., Winship, C. 2002. Commentary: Population versus Individual Level Causal Effects. International Journal of Epidemiology 31, 432–434.

Evans, J. L. 1954. Knowledge and Behavior. Proceedings of the Aristotelian Society 54, 27–48.

Faber, K.-G. 1971. Theorie der Geschichtswissenschaft. München: Beck.

Fisher, J. 1975. Knowledge of Rules. Review of Metaphysics 28, 237–260.

Freedman, D. A. 1992. As Others See Us: A Case Study in Path Analysis. In: J. P. Shaffer (ed.), The Role of Models in Nonexperimental Social Science: Two Debates, 3–30. Washington: American Educational Research Association.

Freedman, D. A. 1997. From Association to Causation via Regression, In: V. McKim, S. P. Turner (eds.), Causality in Crisis? Statistical Methods and the Search for Causal Knowledge in the Social Sciences, 113–161. Notre Dame: University of Notre Dame Press.

Freeman, L. C. 1992. The Sociological Concept of „Group“: An Empirical Test of Two Models. American Journal of Sociology 98, 152–166.

Frey, G. 1961. Symbolische und Ikonische Modelle. In: H. Freudenthal (ed.), The Concept and the Role of the Model in Mathematics and Natural and Social Sciences, 89 – 97. Dordrecht: Reidel.

Fürstenberg, F. 1966. „Sozialstruktur“ als Schlüsselbegriff der Gesellschaftsanalyse. Kölner Zeitschrift für Soziologie und Sozialpsychologie 18, 439–453.

Glenn, N. D. 1977. Cohort Analysis. Beverly Hills: Sage.

Glymour, C. 1999. Preface. In: C. Glymour, G. F. Cooper (eds.), Computation, Causation, and Discovery, xi-xv. Cambridge: MIT Press.

Glymour, C., Spirtes, P., Scheines, R. 1991. Causal Inference. Erkenntnis 35, 151–189.

Goldthorpe, J. H. 2001. Causation, Statistics, and Sociology. European Sociological Review 17, 1–20.

Granger, C. W. J. 1990. Causal Inference. In: J. Eatwell, M. Milgate, P. Newman (eds.), The New Palgrave: Econometrics, 45–49. New York: W. W. Norton.

Greenland, S., Pearl, J., Robins, J. M. 1999. Causal Diagrams for Epidemiologic Research. Epidemiology 10, 37–48.

Gumb, R. D. 1972. Rule-Governed Linguistic Behavior. The Hague: Mouton.

Hacker, P. M. S. 1982. Events and Objects in Space and Time. Mind 91, 1–19.

Hausman, D. M. 1993. Linking Causal and Explanatory Asymmetry. Philosophy of Science 60, 435–451.

Hausman, D. M. 1998. Causal Asymmetries. Cambridge: Cambridge University Press.

Heckman, J. J. 1992. Randomization and Social Policy Evaluation. In: C. F. Manski, I. Garfinkel (eds.), Evaluating Welfare and Training Programs, 201–230. Cambridge: Harvard University Press.

Heckman, J. J. 1996. Comment. Journal of the American Statistical Association 91, 459–462.

Heckman, J. J. 2000. Causal Parameters and Policy Analysis in Economics: A Twentieth Century Retrospective. Quarterly Journal of Economics 115, 45–97.

Heckman, J. J. 2005. The Scientific Model of Causality. Sociological Methodology 35, 1–97.

Hendry, D. F., Richard, J.-F. 1982. On the Formulation of Empirical Models in Dynamic Econometrics. Journal of Econometrics 20, 3 – 33.

Hernes, G. 1976. Structural Change in Social Processes. American Journal of Sociology 82, 513–547.

Hitchcock, C. R. 1993. A Generalized Probabilistic Theory of Causal Relevance. Synthese 97, 335–364.

Holland, P. W. 1986. Statistics and Causal Inference. Journal of the American Statistical Association 81, 945–968.

Holland, P. W. 1986a. Rejoinder. Journal of the American Statistical Association 81, 968–970.

Holland, P. W. 1988. Causal Inference, Path Analysis, and Recursive Structural Equation Models. In: C C. Clogg (ed.), Sociological Methodology 1988, 449–484. San Francisco: Jossey-Bass.

Holland, P. W. 1988a: Comment: Causal Mechanism or Causal Effect: Which is Best for Statistical Science? Statistical Science 3, 186–189.

Holland, P. W. 1993. Which Comes First, Cause or Effect? In: G. Keren, C. Lewis (eds.), A Handbook for Data Analysis in the Behavioral Sciences, 273–282. Hillsdale: Lawrence Erlbaum.

Holland, P. W. 1994. Probabilistic Causality Without Probability. In: P. Humphreys (ed.), Patrick Suppes: Scientific Philosopher, Vol. 1, 257–292. Dordrecht: Kluwer.

Holland, P. W. 2001. The Causal Interpretation of Regression Coefficients. In: M. C. Galavotti, P. Suppes, D. Costantini (eds.), Stochastic Causality, 173–187. Stanford: Center for the Study of Language and Information.

Homans, G. C. 1951. The Human Group. London: Routledge & Kegan Paul.

Homans, G. C. 1961. Social Behavior. Its Elementary Forms. New York: Harcourt, Brace & World.

Homans, G. C. 1976. What Do We Mean by Social „Structure“? In: P. M. Blau (ed.), Approaches to the Study of Social Structure, 53–65. London: Open Books.

International Statistical Institute 1986. Declaration on Professional Ethics. International Statistical Review 54, 227–242.

Jauss, H. R. 1973. Versuch einer Ehrenrettung des Ereignisbegriffs. In: R. Koselleck, W.-D. Stempel (Hg.), Geschichte – Ereignis und Erzählung, 554–560. München: Fink.

Joas, H. 2001. Die soziologische Perspektive. In: Ders. (Hg.), Lehrbuch der Soziologie, 11–38 Frankfurt: Campus.

Joerges, B. 1989. Technische Normen – Soziale Normen? Soziale Welt 40, 242–258.

Kemmerling, A. 1975. Regel und Geltung im Lichte der Analyse Wittgensteins. Rechtstheorie 6, 104–131.

Kendall, M., Stuart, A. 1977. The Advanced Theory of Statistics, Vol. 1 (4th ed.). London: Charles Griffin & Comp.

King, G., Keohane, R. O., Verba, S. 1994. Designing Social Inquiry. Scientific Inference in Qualitative Research. Princeton: Princeton University Press.

Krackhardt, D. 1987. Cognitive Social Structures. Social Networks 9, 109–134.

Krüger, L. 1976. Are Statistical Explanations Possible? Philosophy of Science 43, 129–146.

Leff, G. 1969. History and Social Theory. London: Merlin Press.

Lepsius, M. R. 1976. Zum Verhältnis von Geschichtswissenschaft und Soziologie. In: H. M. Baumgartner, J. Rüsen (Hg.), Seminar: Geschichte und Theorie, 118–138. Frankfurt: Suhrkamp.

Lindenberg, S. 1977. Individuelle Effekte, kollektive Phänomene und das Problem der Transformation. In: K. Eichner, W. Habermehl (Hg.), Probleme der Erklärung sozialen Verhaltens, 46–84. Meisenheim: Hain.

Lübbe, H. 1965. Zur Theorie der Entscheidung. In: Ders., Theorie und Entscheidung. Studien zum Primat der praktischen Vernunft, 7–31. Freiburg: Rombach 1971.

Maddala, G. S. 1977. Self-Selectivity Problems in Econometric Models. In: P. R. Krishnaiah (ed.), Applications of Statistics, 351–366. Amsterdam: North-Holland.

Maddala, G. S. 1983. Limited-dependent and Qualitative Variables in Econometrics. Cambridge: University Press.

Marini, M. M., Singer, B. 1988. Causality in the Social Sciences. C. C. Clogg (ed.), Sociological Methodology 1988, 347–409. San Francisco: Jossey-Bass.

Marsden, P. V., Laumann, E. O. 1984. Mathematical Ideas in Social Structural Analysis. Journal of Mathematical Sociology 10, 271–294.

Marsden, P. V., Lin, N. 1982. Introduction. In: Dies. (Hg.), Social Structure and Network Analysis, 9–11. Beverly Hills: Sage.

McKim, V. R. 1997. Introduction. In: V. R. McKim, S.P. Turner (eds.), Causality in Crisis? Statistical Methods and the Search for Causal Knowledge in the Social Sciences, 1–19. Notre Dame: University of Notre Dame Press.

Meier, C. 1978. Fragen und Thesen zu einer Theorie historischer Prozesse. In: K.-G. Faber, C. Meier (Hg.), Historische Prozesse, 11–66. München: Deutscher Taschenbuch-Verlag.

Meran, J. 1979. Individualismus oder Kollektivismus? Zeitschrift für allgemeine Wissenschaftstheorie 10, 35–53.

Merton, R. K. 1957. Social Theory and Social Structure. Glencoe: Free Press.

Mill, J. S. 1875. Über Religion. Natur. Die Nützlichkeit der Religion. Theismus. Drei nachgelassene Essays. Dt. von E. Lehmann. Berlin: Franz Duncker.

Moscovici, S. 1982. Versuch über die menschliche Geschichte der Natur. Frankfurt: Suhrkamp.

Mueller, U. 1993. Bevölkerungsstatistik und Bevölkerungsdynamik. Berlin: de Gruyter.

Oakeshott, M. 1983. On History and Other Essays. Oxford: Basil Blackwell.

Oaklander, L. N., Smith, Q. (eds.) 1994. The New Theory of Time. New Haven: Yale University Press.

Olkin, I., Gleser, L. J., Derman, C. 1980. Probability Models and Applications. New York: Macmillan Publ.

O'Neill, J. (ed.) 1973. Modes of Individualism and Collectivism. London: Heinemann.

Nurmi, H. 1974. Social Causality and Empirical Data Reduction Techniques. Quality and Quantity 8, 159–180.

Opp, K.-D. 1983. Die Entstehung sozialer Normen. Tübingen: Mohr-Siebeck.

Opp, K.-D., Hummell, H. J. 1973. Soziales Verhalten und soziale Systeme. Frankfurt: Athenäum.

Oppitz, M. 1975. Notwendige Beziehungen. Abriß der strukturalen Anthropologie. Frankfurt: Suhrkamp.

Ort, C.-M. 2003. Kulturbegriffe und Kulturtheorien. In: A. Nünning, V. Nünning (Hg.), Konzepte der Kulturwissenschaften, 19–38. Stuttgart: J. B. Metzler.

Papineau, D. 2001. Metaphysics over Methodology – or, Why Infidelity Provides No Grounds to Divorce Causes from Probabilities. In: M. C. Galavotti, P. Suppes, D. Costantini (eds.), Stochastic Causality, 15–38. Stanford: Center for the Study of Language and Information.

Pappi, F. U. 1987. Die Netwerkanalyse aus soziologischer Perspektive. In: Ders. (Hg.), Methoden der Netzwerkanalyse, 11–37. München: Oldenbourg.

Pearl, J. 1998. Graphs, Causality, and Structural Equation Models. Sociological Methods & Research 27, 226–284.

Pearl, J. 1999. Graphs, Structural Models, and Causality. In: C. Glymour, G. F. Cooper (eds.), Computation, Causation, and Discovery, 95-138. Cambridge: MIT Press.

Pearl, J. 2000. Causality. Cambridge: Cambridge University Press.

Pearl, J. 2001. Causal Inference in the Health Sciences: A Conceptual Introduction. Health Services & Outcomes Research Methodology 2, 189-2001.

Peterson, R. A. 1979. Revitalizing the Culture Concept. Annual Review of Sociology 5, 137–166.

Pötter, U. 2006. Probabilistische Selektionsmodelle. In: A. Diekmann (Hg.), Methoden der Sozialforschung, 172-202. Wiesbaden: VS Verlag für Sozialwissenschaften.

Pratt, J. W., Schlaifer, R. 1984. On the Nature and Discovery of Structure. Journal of the American Statistical Association 79, 9–33.

Prendergast, C., Knottnerus, J. D. 1994. Recent Developments in the Theory of Social Structure: Introduction and Overview. In: J. D. Knottnerus, C. Prendergast (eds.), Current Perspectives in Social Theory, Suppl. 1, 1–26. Greenwich: JAI Press.

Pressat, R. 1972. Demographic Analysis. Methods, Results, Applications. Transl. from French by J. Matras. Foreword by N. Keyfitz. Chicago: Aldine & Atherton.

Quine, W. V. 1972. Methodological Reflections on Current Linguistic Theory. In: D. Davidson, G. Harman (eds.), Semantics of Natural Language, 442–454. Dordrecht: Reidel.

Reinhold, G. (Hg.) 2000. Soziologie-Lexikon. 4. Aufl. München: Oldenbourg.

Richards, E. G. 1998. Mapping Time. The Calendar and its History. Oxford: Oxford University Press.

Robins, J. M. 1999. Association, Causation, and Marginal Structural Models. Synthese 121, 151–179.

Rohwer, G., Pötter, U. 2001. Grundzüge der sozialwissenschaftlichen Statistik. Weinheim: Juventa.

Rohwer, G., Pötter, U. 2002a. Methoden sozialwissenschaftlicher Datenkonstruktion. Weinheim: Juventa.

Rohwer, G., Pötter, U. 2002b. Wahrscheinlichkeit. Begriff und Rhetorik in der Sozialforschung. Weinheim: Juventa.

Rosenbaum, P. R. 1984. From Association to Causation in Observational Studies: The Role of Tests of Strongly Ignorable Treatment Assignment. Journal of the American Statistical Association 79, 41–48.

Rosenbaum, P. R. 1999. Choice as an Alternative to Control in Observational Studies. Statistical Science 14, 259–304.

Rosenbaum, P. R. 2002. Observational Studies. 2nd Ed. New York: Springer.

Rosenbaum, P. R., Rubin, D. B. 1983. The Central Role of the Propensity Score in Observational Studies for Causal Effects. Biometrika 70, 41–55.

Rosenbaum, P. R., Rubin, D. B. 1984. Comment: Estimating the Effects Caused by Treatments. Journal of the American Statistical Association 79, 26–28.

Rothman, K. J., Greenland, S. 2005. Causation and Causal Inference in Epidemiology. American Journal of Public Health 95, Supp. 1, 144–150.

Röttgers, K. 1983. Der Ursprung der Prozessidee aus dem Geiste der Chemie. Archiv für Begriffsgeschichte 27, 93–157.

Rubin, D. B. 1974. Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. Journal of Educational Psychology 66, 688–701.

Rubin, D. B. 1977. Assignment to Treatment Group on the Basis of a Covariate. Journal of Educational Statistics 2, 1–26.

Rubin, D. B. 1978. Bayesian Inference for Causal Effects: The Role of Randomization. Annals of Statistics 6, 34–58.

Rubin, D. B. 1986. Comment: Which Ifs Have Causal Answers. Journal of the American Statistical Association 81, 961–962.

Rubin, D.B. 1990. Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies. Statistical Science 5, 472–480.

Ryle, G. 1949. The Concept of Mind. Dt.: Der Begriff des Geistes. Stuttgart: Reclam 1982.

Scheines, R. 1997. An Introduction to Causal Inference. In: V. R. McKim, S. P. Turner (eds.), Causality in Crisis? Statistical Methods and the Search for Causal Knowledge in the Social Sciences, 185–199. Notre Dame: University of Notre Dame Press.

Scheuch, E. K., Daheim, H. 1961. Sozialprestige und soziale Schichtung. In: D. V. Glass, R. König (Hg.), Soziale Schichtung und soziale Mobilität, 65–103. Opladen: Westdeutscher Verlag.

Scheuch, E. K., Kutsch, T. 1975. Grundbegriffe der Soziologie, Band 1: Grundlegung und elementare Phänomene. 2. Aufl. Stuttgart: Teubner.

Schütz, A. 1991. Der sinnhafte Aufbau der sozialen Welt. 5. Aufl. Frankfurt: Suhrkamp.

Searle, J. R. 1969. Speech Acts. An Essay in the Philosophy of Language. Cambridge: Cambridge University Press.

Searle, J. R. 1987. Intentionalität. Frankfurt: Suhrkamp.

Sewell, W. H. 1992. A Theory of Structure: Duality, Agency, and Transformation. American Journal of Sociology 98, 1–29.

Shadish, W. R., Cook, T. D., Campbell, D. T. 2002. Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Boston: Houghton Mifflin.

Smith, B. C. 1996. On the Origin of Objects. Cambridge: MIT Press.

Sobel, M. E. 1995. Causal Inference in the Social and Behavioral Sciences. In: G. Arminger, C. C. Clogg, M. E. Sobel (eds.), Handbook of Statistical Modeling for the Social and Behavioral Sciences, 1–38. New York: Plenum Press.

Sobel, M. E. 2005. Discussion: 'The Scientific Model of Causality'. Sociological Methodology 35, 99–133.

Sober, E. 1986. Causal Factors, Causal Inference, Causal Explanation. The Aristotelian Society. Suppl. Vol. 60, 97–113.

Soper, K. 1995. What is Nature? Culture, Politics and the non-Human. Oxford: Blackwell.

Spirtes, P., Glymour, C., Scheines, R. 1993. Causation, Prediction, and Search. New York: Springer.

Splawa-Neyman, J. 1990. On the Application of Probability Theory to Agricultural Experiments. Statistical Science 5, 465–480.

Stegmüller, W. 1983. Erklärung, Begründung, Kausalität (Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie, Band I). Berlin: Springer-Verlag.

Steiner, H.-G. 1969. Aus der Geschichte des Funktionsbegriffs. Der Mathematikunterricht 15, 13–39.

Steyer, R., Gabler, S., Davier, A. A. von, Nachtigall, C., Buhl, T. 2000. Causal Regression Models I: Individual and Average Causal Effects. Methods of Psychological Reasearch Online 5, No. 2, 39–71. `http://www.pabst-publishers.de/mpr/`

Suppes, P. 1970. A Probabilistic Theory of Causality. Amsterdam: North-Holland.

Suppes, P., Zanotti, M. 1996. Foundations of Probability with Applications. Selected Papers, 1974–1995. Cambridge: Cambridge University Press.

Susser, M. 1991. What is a Cause and How do we Know One? A Grammar for Pragmatic Epidemiology. American Journal of Epidemiology 133, 635–648.

Tryfos, P. 2004. The Measurement of Economic Relationships. Dordrecht: Kluwer.

Vanberg, V. 1975. Die zwei Soziologien. Individualismus und Kollektivismus in der Sozialtheorie. Tübingen: Mohr.

Wasserman, S., Faust, K. 1994. Social Network Analysis: Methods and Applications. Cambridge: Cambridge University Press.

Weber, M. 1921. Wirtschaft und Gesellschaft. 1. Halbband, 5. Aufl., hrsg. von J. Winckelmann. Tübingen: Mohr-Siebeck 1976.

Weymann, A. 2001. Interaktion, Sozialstruktur und Gesellschaft. In: H. Joas (Hg.), Lehrbuch der Soziologie, 93–121. Frankfurt: Campus.

Winship, C., Morgan, S. L. 1999. The Estimation of Causal Effects from Observational Data. Annual Review of Sociology 25, 659–706.

Wirth, E. 1979. Theoretische Geographie. Grundzüge einer Theoretischen Kulturgeographie. Stuttgart: Teubner.

Wittgenstein, L. 1921. Tractatus logicophilosophicus. Frankfurt: Edition Suhrkamp 1963.

Woodward, J. 1999. Causal Interpretation in Systems of Equations. Synthese 121, 199–247.

Woodward, J. 2001. Probabilistic Causality, Direct Causes and Counterfactual Dependence. In: M. C. Galavotti, P. Suppes, D. Costantini (eds.), Stochastic Causality, 39–63. Stanford: Center for the Study of Language and Information.

Woodward, J. 2003. Making Things Happen. A Theory of Causal Explanation. Oxford: Oxford University Press.

Wright, G. H. von 1971. Explanation and Understanding. New York: Cornell University Press. Dt.: Erklären und Verstehen. Königstein: Athenäum 1984.

Wright, G. H. v. 1978. Erwiderungen. In: K. O. Apel, J. Manninen, R. Tuomela (Hg.), Neue Versuche über Erklären und Verstehen, 264–302. Frankfurt: Suhrkamp.

Zapf, W. 1995. Entwicklung und Sozialstruktur moderner Gesellschaften. In: H. Korte, B. Schäfers (Hg.), Einführung in Hauptbegriffe der Soziologie, 181–193. Opladen: Leske + Budrich.

Namenverzeichnis

Stichwortverzeichnis