Models in Statistical Social Research

Götz Rohwer

ROUTLEDGE

Social Research Today

Models in Statistical Social Research

Götz Rohwer

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

First published 2010
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

Simultaneously published in the USA and Canada
by Routledge
270 Madison Avenue, New York, NY 10016

Routledge is an imprint of the Taylor & Francis Group, an informa business

This edition published in the Taylor & Francis e-Library, 2009.

To purchase your own copy of this or any of Taylor & Francis or Routledge's collection of thousands of eBooks please go to www.eBookstore.tandf.co.uk

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging in Publication Data
Rohwer, Götz.
Models in statistical social research / Götz Rohwer.
p. cm.
Includes bibliographical references.
1. Social sciences–Statistical methods. I. Title.
HA29.R73 2010
300.72'7–dc22 2009024400

ISBN 0-203-86580-4 Master e-book ISBN

ISBN10: 0-415-56055-1 (hbk)
ISBN10: 0-203-86580-0 (ebk)

ISBN13: 978-0-415-56055-9 (hbk)
ISBN13: 978-0-203-86580-4 (ebk)

Contents

List of figures ix
List of tables x
Preface xi
List of symbols xiv

1 Variables and relations 1
1.1 Variables and distributions 1
1.2 Relations 11

2 Notions of structure 16
2.1 Statistical notions of structure 17
2.2 Taking relations into account 22

3 Processes and process frames 26
3.1 Historical and repeatable processes 26
3.2 Time series and statistical processes 30
3.3 Stochastic process frames 32

4 Functional models 36
4.1 Deterministic models 38
4.2 Models with stochastic variables 44
4.3 Exogenous and unobserved variables 50

5 Functional causality 56
5.1 Functional causes and conditions 57
5.2 Ambiguous references to individuals 61
5.3 Isolating functional causes 65

6 Models and statistical data 70

6.1 Functional models and data 71

6.2 Experimental and observational data 74

6.3 Interventions and reference problems 82

7 Models with event variables 93

7.1 Situations and events 94

7.2 Event models with time axes 99

7.3 Dynamic causality 104

8 Multilevel and population-level models 111

8.1 Conceptual frameworks 112

8.2 Models of statistical processes 119

8.3 Functional causality and levels 123

Notes 130

References 138

Index 144

List of figures

1.1 Graphical presentation of a frequency distribution 6
1.2 Illustration of a regression function 10
4.1 Dependency of $\Pr(\dot{X} = 1 | \dot{Y} = 1)$ on $\Pr(\dot{X} = 1)$ 50
6.1 Dependency of causal effects on statistical distributions 86
8.1 Selection probabilities and statistical distributions 119
8.2 Ten simulated diffusion processes 121
8.3 Individual-level effects of changing distributions 124
8.4 Population-level effects of changing distributions 125
8.5 Illustration of self selection with constraints 127
8.6 A structured diffusion process with two groups 128
8.7 Group-specific diffusion rates 128
8.8 Group-specific diffusion rates 128
8.9 Overall diffusion rates 129

List of tables

1.1 Fictitious data for a statistical variable 3
1.2 Age distribution in Germany 1999 5
1.3 Household income and expenditures on food 9
1.4 An example of network data 12
1.5 Adjacency matrix of a relational variable 13

Preface

This book is about models in statistical social research. The expression 'statistical social research' is here used to denote approaches that start from formally defined variables (in contrast, e.g. to case studies); individual cases only exemplify the variables, and their identities are ignored.

The models constructed in these approaches to social research may serve quite different purposes. A basic distinction is between descriptive and analytical models. Descriptive models serve to represent something and thereby provide (in some sense) a more useful view of that thing. In the realm of statistical social research these are most often models representing statistical distributions, or certain features of such distributions, as they happen to exist in some specified populations. Analytical models, on the other hand, do not intend to provide descriptions but permit one to deal with modal questions and, in particular, with causal and non-causal dependency relations between variables.

The distinction between factual and modal questions (and notions) is fundamental for much of the argument in the present text. Factual questions relate to facts and, from a temporal view, are retrospective questions referring to what happened in the past. This is particularly true of statistical descriptions which conceptually presuppose a data generating process. In contrast, modal questions refer to possibilities and to modalities which relate to possibilities (like probabilities). An important kind of modal question concerns dependency relations between events and/or states of affairs.

Corresponding to the distinction between factual and modal questions, a distinction between two kinds of variables is proposed. On the one hand, there are variables intended to represent (actually observed or hypothetically assumed) facts. They are called *statistical variables* because they represent statistical data. On the other hand, *modal variables* serve to formulate modal questions and hypotheses (e.g. questions concerning probabilities in a hypothetical experiment).

As will be discussed in Chapter 1, statistical variables provide a useful starting point for a systematic development of all basic concepts of descriptive statistics (broadly understood as statistical methods to be used for descriptive purposes). This chapter also introduces an extension of the statistical framework that allows an explicit representation of relations. Based on these frameworks, the second chapter argues that there is no clearcut opposition between statistical and relational notions of structure, and several possibilities exist for the statistical approach to take into account relational properties. Limitations only result from the presupposition that individual units cannot be identified.

The main interest of the present text concerns analytical models. In order to approach an understanding, we begin (in Chapter 3) with a distinction between historical processes, understood as unique developments in historical time, and process frames to be used as conceptual frameworks for the definition of (actually or hypothetically) repeatable processes. Given this distinction, a broad class of analytical models can be understood as relating to process frames for repeatable processes.

This book argues that such models can best be understood as *functional models*, models that use deterministic or stochastic functions to connect modal variables. As will be discussed in Chapters 5 and 6, these models also provide a useful framework for a definition of "causal relationships." Our proposal is based on a distinction between values of variables that can be viewed as *conditions* of values of other variables and changes of values of variables that can be viewed as *causes* of changes of values of other variables. Since the definitions relate to functional models this approach is termed *functional causality*.

Stochastic functional models proposed in the statistical literature are often conceptualized as *data models*, that is, models relating to a data generating process assumed for a given set of data. This approach allows one to think of the model's exogenous variables as stochastic variables having a given distribution. However, data generating processes must be distinguished from substantial processes that actually generate events and states for which afterwards data might become available. If to be used as analytical models, functional models must relate to a process frame for substantial processes without providing an empirical basis for assumptions about exogenous variables. In contrast to data models, analytical functional models for substantial processes are then theoretical constructs without a direct relationship to the empirical world.

The distinction between substantial and data generating processes also helps to understand randomized experiments, namely as process frames for substantial processes. Chapter 6 develops the argument that the conceptual framework of randomized experiments is of only limited use in social research and argues that (self) selection processes should not be understood as biasing "true" causal relationships but as being a substantial part of the social processes to be modeled.

Several possibilities exist for the formulation of functional models for substantial processes. A first alternative concerns the representation of time. One can use an implicit representation of time by interpreting the functions of a functional model as implying a temporal relationship. Alternatively, one can use a time axis to define processes as sequences of variables (e.g. time series or statistical processes defined as sequences of statistical variables).

A second alternative concerns whether processes are defined in terms of states or events. Time series and statistical processes are almost always defined in terms of state variables (variables recording the possibly changing state of something).

Alternatively, one can think of processes as temporally structured series of events. A formal representation requires event variables which are conceptually different from state variables. Within this framework, Chapter 7 introduces a dynamic (as opposed to only comparative) version of functional causality.

A third alternative concerns whether a functional model relates to a generic individual unit or to a collection of individual units. Statistical social research is predominantly occupied with individual-level models. Even then it may be necessary to use multilevel models that include population-level variables. However, important questions also concern relationships between statistical distributions (defined for some population). Such questions cannot, in general, be answered from individual-level models. They require population-level models which allow taking into account constraints and interdependencies at the population level. Such models will be discussed in Chapter 8.

I thank Ulrich Pötter for many helpful discussions and also acknowledge that many ideas developed in the present text derive from earlier work done together with him (see the references). For helpful comments I also thank Christian Dudel, Sebastian Jeworutzki, and Bernhard Schimpl-Neimanns. Lastly, I would like to thank John Haisken-DeNew for his help in polishing the English used in this book.

G. Rohwer
Bochum, June 2009

List of symbols

The table shows the special symbols that are most often used. Brackets indicate sections where the notations are explained.

$X, Y, Z, \ldots$	statistical variables (1.1)
Ω	reference set of a statistical variable (1.1)
$\dot{X}, \dot{Y}, \dot{Z}, \ldots$	stochastic modal variables (3.3, 4.2)
$\ddot{X}, \ddot{Y}, \ddot{Z}, \ldots$	deterministic modal variables (3.2, 4.1)
$\tilde{\mathcal{X}}, \tilde{\mathcal{Y}}, \tilde{\mathcal{Z}}, \ldots$	property spaces of statistical or modal variables (1.1)
$\tilde{x}, \tilde{y}, \tilde{z}, \ldots$	property values (elements of property spaces)
$\tilde{X}, \tilde{Y}, \tilde{Z}, \ldots$	property sets (subsets of property spaces)
$\mathrm{P}[X]$	frequency distribution of the statistical variable X (1.1)
$\mathrm{Pr}[\dot{X}]$	probability distribution of the stochastic variable $\dot{X}$ (3.3)
$\mathrm{M}(X)$	mean of the statistical variable X
$\mathrm{E}(\dot{X})$	mean (expectation) of the stochastic variable $\dot{X}$
$\Delta(x', x'')$	change of a modal variable from x' to x'' (5.1)
$\dot{E}, \dot{E}_1, \dot{E}_2, \ldots$	stochastic event variables (7.1)
$\ddot{E}, \ddot{E}_1, \ddot{E}_2, \ldots$	deterministic event variables (7.1)
$\tilde{\mathcal{E}}, \tilde{\mathcal{E}}_1, \tilde{\mathcal{E}}_2, \ldots$	ranges of event variables (7.1)
$\mathbf{R}$	set of real numbers
$\lvert M \rvert$	number of elements of a finite set M
$\mathcal{P}(M)$	set of all subsets (power set) of M

1 Variables and relations

The text is partitioned into chapters and sections as indicated in the table of contents. A further subdivision into paragraphs (denoted by §) is provided at the beginning of each chapter.

1.1 Variables and distributions
1. Statistical variables
2. A fictitious illustration
3. Multidimensional variables
4. Statistical distributions
5. Descriptive statistical statements
6. Conditional distributions
7. Regression functions
8. Descriptive regression models
9. Statistical and substantial conditions

1.2 Relations
1. Relational variables
2. Construction of networks
3. Formal descriptions of networks
4. Different kinds of relations
5. Factual and modal views of relations

The first section introduces elementary statistical concepts for descriptive purposes, in particular, statistical variables and distributions, conditional distributions, and regression functions. The second section extends the statistical framework to allow for an explicit formal representation of relations and then discusses different notions of relations.

1.1 Variables and distributions

1. Statistical variables

Most statistical concepts derive from the notion of a statistical variable. Unfortunately, the word ‘variable’ is easily misleading because it suggests something that "varies" or being a "variable quantity."[1] In order to get an appropriate understanding it is first of all necessary to distinguish statistical from logical variables. Consider the expression '$x \leq 5$'. In this expression, x is a *logical variable* that can be replaced by a name. Obviously, without substituting a specific name, the expression '$x \leq 5$' has no definite meaning and, in particular, is neither a true nor a false statement. The expression is actually no statement at all but a *sentential function*. A statement that is true or false or meaningless only results when a name is substituted for x. For example, if the symbol 1 is substituted for x, the result is a true statement ($1 \leq 5$); if the symbol 9 is substituted for x, the result is a false statement ($9 \leq 5$); and if some name not referring to a number is substituted for x, the result is meaningless. Such logical variables are used, for example, in mathematics to formulate general statements. Statistical variables serve a quite different purpose. They are used to represent the data for statistical calculations which refer to properties of objects. The basic idea is that one can characterize objects by properties. Since this is essentially an assignment of properties to objects, statistical variables are defined as functions:[2]

$$X : \Omega \longrightarrow \tilde{\mathcal{X}}$$

X is the name of the function, Ω is its domain, and $\tilde{\mathcal{X}}$ is the codomain (a set of possible values), also called the *range* of X. To each element $\omega \in \Omega$, the statistical variable X assigns exactly one element of $\tilde{\mathcal{X}}$ denoted by $X(\omega)$. In this sense, a statistical variable is simply a function.[3] What distinguishes statistical variables from other functions is a specific purpose: statistical variables serve to characterize objects. Therefore, in order to call X a statistical variable (and not just a function), its domain, Ω, should be a set of objects and its codomain, $\tilde{\mathcal{X}}$, should be a set of properties that can be meaningfully used to characterize the elements of Ω. To remind of this purpose, the set of possible values of a statistical variable will be called its *characteristic* or *property space* and its elements will be called *property values*.[4]

In the statistical literature domains of statistical variables are often called populations. This is unfortunate because a statistical variable can refer to any kind of object. We therefore often prefer to speak of the *domain* or, equivalently, the *reference set* of a statistical variable.

2. A fictitious illustration

A simple example can serve to illustrate the notion of a statistical variable. In this example, the reference set is a set of 10 people, symbolically $\Omega := \{\omega_1, \ldots, \omega_{10}\}$.[5] The variable, denoted X, is intended to represent, for each member of Ω, the sex. This can be done with a property space $\tilde{\mathcal{X}} := \{0, 1\}$, with elements 0 (meaning 'male') and 1 (meaning 'female'). Then, for each member $\omega \in \Omega$, $X(\omega)$ is a value in $\tilde{\mathcal{X}}$ and shows ω's sex.

Of course, in order to make use of a statistical variable one needs data. In contrast to most functions that are used in mathematics, statistical variables cannot

Table 1.1 Fictitious data for a statistical variable X (left-hand side) and a two-dimensional statistical variable (X, Y) (right-hand side)

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

be defined by referring to some kind of rule. There is no rule that allows one to infer of the sex, or any other property, of an individual by knowing its name. In order to make a statistical variable explicitly known one almost always needs a tabulation of its values. The left-hand side of Table 1.1 provides fictitious data as an illustration for the current example.

Note that it is general practice in statistics to *represent* properties (the elements of a property space) by numbers. One reason for doing so is the resulting simplification in the tabulation of statistical data. The main reason is, however, another one: numerical representations allow the performing of statistical calculations. The *mean value* of a statistical variable can serve as an example. The definition is $\mathrm{M}(X) := \sum_{\omega \in \Omega} X(\omega)/|\Omega|$. The calculation consists in summing up the values of the variable for all elements in the reference set and then dividing by the number of elements.[6] Obviously, the calculation requires a numerical representation for the values of the variable, that is, for the elements of its property space. But as soon as one has introduced a numerical representation one can do anything that can be done with numbers too with the values of a variable. To be sure, this does not guarantee a result with an immediate and sensible interpretation. This might, or might not, be the case and can never be guaranteed from a statistical calculation alone.[7] However, in the present example one gets a sensible result. Performing the calculation of a mean value for the variable X, the result is $\mathrm{M}(X) = 0.4$, providing the proportion of female individuals in the reference set.

3. *Multidimensional variables*

It is often possible and informative to characterize objects simultaneously by several properties. In the example of the previous paragraph one can assume that each person can also be assigned a specific age. Correspondingly, one uses a two-dimensional variable

$$(X, Y) : \Omega \longrightarrow \tilde{\mathcal{X}} \times \tilde{\mathcal{Y}}$$

which combines two property spaces: $\tilde{\mathcal{X}}$ for the sex and another property space $\tilde{\mathcal{Y}} := \{0, 1, 2, \ldots\}$ for age (recorded, e.g. in completed years). The right-hand side of Table 1.1 provides again fictitious data. For example, $(X, Y)(\omega_1) = (0, 22)$, meaning that ω_1 is the name of a man of age 22.

Obviously, there can be any number of dimensions. The expression has no spatial meaning and simply refers to a property space in the sense of a set of attributes. In fact, the number of dimensions has no substantial meaning at all because two or more dimensions can always be combined into a single property space. Conversely, referring to a formally one-dimensional variable, say $X : \Omega \longrightarrow \tilde{\mathcal{X}}$, one can always assume that it is a short-hand notation for several dimensions, e.g. $X = (X_1, \ldots, X_m)$ with a property space $\tilde{\mathcal{X}} = \tilde{\mathcal{X}}_1 \times \cdots \times \tilde{\mathcal{X}}_m$.

4. Statistical distributions

The notion of statistical variables provides a quite general conceptual framework for the representation of objects (of any kind) and their properties. Knowing a variable $X : \Omega \longrightarrow \tilde{\mathcal{X}}$, one also knows for each element $\omega \in \Omega$ the assigned property value $X(\omega)$. However, the statistical concern is not with the individual members of Ω. For example, referring to Table 1.1, it is of no interest that ω_1 is the name of a male and ω_2 is the name of a female person; the statistically relevant information rather is that Ω consists of 60 per cent male and 40 per cent female persons.[8] This is then called the *statistical distribution* of X (in the reference set Ω). Thus, from a statistical point of view, statistical variables only serve to derive statistical distributions. Referring to a variable $X : \Omega \longrightarrow \tilde{\mathcal{X}}$, its statistical distribution is a function which provides, for each subset $\tilde{X} \subseteq \tilde{\mathcal{X}}$, a number

$$\mathrm{P}(X \in \tilde{X}) := \frac{|\{\omega \in \Omega \,|\, X(\omega) \in \tilde{X}\}|}{|\Omega|} \tag{1.1}$$

It is the proportion of members of Ω which are assigned by the variable X a property value in the set $\tilde{X}$. If $\tilde{X}$ consists of a single element, say x, the notation $\mathrm{P}(X = x)$ is also used. Using, for example, the figures from Table 1.1, $\mathrm{P}(X = 0) = 0.6$, $\mathrm{P}(Y = 25) = 0.3$, $\mathrm{P}(Y \in \{22, 23, 24\}) = 0.4$. In the last case, an equivalent notation would be $\mathrm{P}(Y \leq 24)$.

Like statistical variables, statistical distributions are also functions. However, the domain is no longer the original reference set of objects but the power set of a property space. In a formal notation, the distribution of a variable $X : \Omega \longrightarrow \tilde{\mathcal{X}}$, also called its *frequency function*, is a function

$$\mathrm{P}[X] : \mathcal{P}(\tilde{\mathcal{X}}) \longrightarrow \mathbf{R} \tag{1.2}$$

providing, for each element $\tilde{X} \in \mathcal{P}(\tilde{\mathcal{X}})$, the number $\mathrm{P}[X](\tilde{X}) := \mathrm{P}(X \in \tilde{X})$ as defined in (1.1).[9] The notation $\mathrm{P}[X]$ is used when referring to the distribution of a statistical variable in a general sense. In order to indicate to which variable the distribution refers, its name is appended in square brackets. These square brackets form a part

Table 1.2 Number of men (n_τ^m) and women (n_τ^f) of age τ in the year 1999 in Germany; 95* refers to all ages $\tau \geq 95$

τ	n_τ^m	n_τ^f	τ	n_τ^m	n_τ^f	τ	n_τ^m	n_τ^f
0	399633	378251	32	732651	686810	64	472336	511058
1	410782	389437	33	748125	697478	65	409837	445060
2	413836	391872	34	758218	705906	66	361225	399667
3	403107	381972	35	761731	711736	67	359314	405178
4	398813	377741	36	746559	700436	68	365748	423645
5	409761	387869	37	727433	687469	69	362427	431229
6	422128	400905	38	711239	675052	70	347956	425406
7	435168	413392	39	690941	656171	71	319299	416670
8	466447	442107	40	663561	630073	72	283797	410392
9	485976	461036	41	641087	608883	73	258704	416006
10	490076	464833	42	627205	597762	74	228253	407353
11	492537	465361	43	610005	584657	75	202335	390714
12	482637	456705	44	594357	575662	76	196282	387353
13	469636	445784	45	578806	566937	77	193182	395011
14	461216	437317	46	570259	561714	78	180053	390951
15	463159	438441	47	566062	558353	79	144258	328396
16	472798	446710	48	563942	556038	80	96563	221190
17	479914	453245	49	558611	548502	81	68960	161774
18	481413	457174	50	529900	517217	82	65068	156443
19	473334	451301	51	494176	482985	83	70399	177521
20	462189	441633	52	450618	442241	84	79463	213228
21	460967	441747	53	397251	392979	85	78125	219593
22	460272	441953	54	434885	432973	86	67852	197433
23	456346	436715	55	502194	498201	87	55469	168730
24	458828	438622	56	500889	496426	88	44330	141995
25	469223	449325	57	545671	544075	89	35770	120032
26	500605	477112	58	612495	615985	90	27775	97204
27	557051	528212	59	621650	630333	91	20817	75582
28	603444	568797	60	593275	606556	92	15772	58386
29	644108	604417	61	551237	569910	93	11724	42726
30	686013	642576	62	522738	546881	94	8658	30973
31	712393	668147	63	502833	534670	95*	21807	69931
						Totals	40047972	42038610

Source: Segment 685 of the STATIS data base of the Statistisches Bundesamt.

of the function name and must not be confused with possible arguments to be appended in round brackets.[10] Note also that identical definitions can be used if X is a multidimensional variable consisting of two or more components.

5. *Descriptive statistical statements*

If statistical statements intend to describe something this is the reference set of a statistical variable. More specifically, given a variable $X : \Omega \longrightarrow \tilde{\mathcal{X}}$, a descriptive statistical statement uses the distribution $\mathrm{P}[X]$ as a description of the reference set Ω. The actual formulation can be in many different ways. Often it is possible to

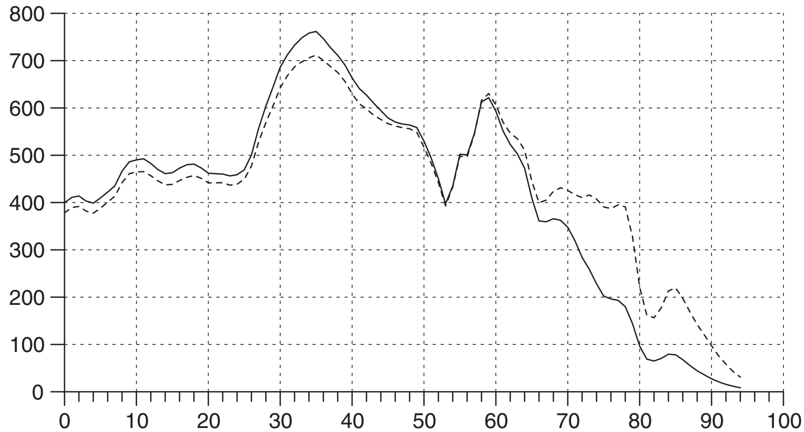


Figure 1.1 Graphical presentation of the absolute frequencies (in 1000) from Table 1.2 until an age of 94 years for men (solid line) and women (dotted line).

present the complete distribution in the form of a table or graphically. In addition, or alternatively, one can use quantities derived from statistical distributions, e.g. mean values, quantiles, and measures of dispersion.

As an example, the age distributions of men and women who lived in Germany in the year 1999 will be considered. Two statistical variables, $Y^m : \Omega^m \longrightarrow \tilde{\mathcal{Y}}$ and $Y^f : \Omega^f \longrightarrow \tilde{\mathcal{Y}}$, provide the formal framework. Ω^m represents the male and Ω^f represents the female persons who lived in Germany in the year 1999. The common property space is $\tilde{\mathcal{Y}} := \{0, 1, 2, \ldots\}$ with elements representing age in completed years. Table 1.2 shows the data published by the Statistisches Bundesamt. Obviously, this is already a statistical distribution: for each age the table provides the number of men and the number of women of that age.

The table also illustrates a problem that has motivated the development of many statistical methods: that it is difficult to make sense of the data if simply presented as a tabulated frequency distribution. A famous remark of the statistician R. A. Fisher refers to this problem:

> Briefly, and in its most concrete form, the object of statistical methods is the reduction of data. A quantity of data, which usually by its mere bulk is incapable of entering the mind, is to be replaced by relatively few quantities which shall adequately represent the whole, or which, in other words, shall contain as much as possible, ideally the whole, of the relevant information contained in the original data.
>
> (Fisher 1922: 311)

In the present example one can use a graphical presentation as shown in Figure 1.1 to support the interpretation of the data. One can easily compare male and female age distributions; and the figure also allows the relating of the frequencies to broader age classes and to think of relations with birth frequencies in corresponding birth years.

6. Conditional distributions

Many statistical methods employ the notion of a conditional distribution. This text uses the general notation

$$\mathrm{P}[\text{variables} \,|\, \text{conditions}]$$

in order to denote the distribution of the variables (given before the | sign) restricted to a reference set demarcated by the conditions (given behind the | sign). Like any other distributions, conditional distributions are functions; specific values (frequencies) only result if an argument is appended (in round brackets). To illustrate the notation, the following examples refer to the two-dimensional variable $(X, Y): \Omega \longrightarrow \tilde{\mathcal{X}} \times \tilde{\mathcal{Y}}$ that was introduced in § 3 (see Table 1.1).

(a) $\mathrm{P}[Y|X = 0]$. This is the distribution of Y in the subset $\{\omega|X(\omega) = 0\}$. It is a function, the age distribution of men; specific values result by providing arguments, for example

$$\mathrm{P}(Y \geq 25 \,|\, X = 0) = \mathrm{P}[Y|X = 0](Y \geq 25) = 0.5$$

meaning that 50 per cent of the men are of age 25 or older.[11]

(b) $\mathrm{P}[X\,|\,Y \geq 25]$ is the distribution of X in the set of persons who are of age 25 or older. A specific value is

$$\mathrm{P}(X = 1 \,|\, Y \geq 25) = \mathrm{P}[X\,|\,Y \geq 25](1) = 0.5$$

meaning that 50 per cent of the persons of age 25 or older are female.

(c) $\mathrm{P}[Y\,|\,Y \geq 25, X = 1]$ is the age distribution in the set of women who are of age 25 or older. This example shows that the same variables can be used before and behind the | sign.

7. Regression functions

Often used are methods of regression analysis. In a general sense, these are methods to calculate, and present, conditional distributions. To understand the basic approach, a distinction between general and specific regression functions is useful. Definitions refer to a two-dimensional variable $(X, Y): \Omega \longrightarrow \tilde{\mathcal{X}} \times \tilde{\mathcal{Y}}$. One component, say X, is selected as an *independent*, the other component, Y, as a *dependent* variable. A *general regression function* is a function that assigns to each value $x \in \tilde{\mathcal{X}}$ the conditional frequency function $\mathrm{P}[Y|X = x]$.

The idea is that regression functions can show how values of a variable depend on values of another variable. However, the values of a general regression function are not simple values (numbers) but again functions (distributions). The following diagram illustrates this:

$$\underbrace{x \longrightarrow \overbrace{(y \longrightarrow \mathrm{P}[Y|X=x](y))}^{\text{conditional distribution}}}_{\text{general regression function}}$$

To each value x of the independent variable X, the general regression function assigns a conditional distribution $\mathrm{P}[Y|X=x]$ that is again a function that assigns to each value y of the dependent variable Y the conditional frequency $\mathrm{P}(Y=y\,|\,X=x)$.

This leads to the question how to describe general regression functions. Two (to some extent overlapping) approaches exist. The first approach uses parameterized frequency functions, say $f(y;\theta)$,[12] as a general framework for the distribution of the dependent variable, $\mathrm{P}[Y]$. Then, in order to approximate the conditional distributions, one uses a link function, say $\theta = g(x;\beta)$, that allows the approximation

$$\mathrm{P}(Y=y\,|\,X=x) = \mathrm{P}[Y|X=x](y) \approx f(y; g(x;\beta)) \tag{1.3}$$

The other approach uses *specific regression functions* having the form

$$x \longrightarrow \text{Characterization of } \mathrm{P}[Y|X=x]$$

Values of these functions are numbers which characterize the conditional distributions in some way. Many different possibilities exist. Of widespread use are the following versions.

(a) Conditional mean values. Specific regression functions have then the form $x \longrightarrow \mathrm{M}(Y|X=x)$. The function assigns to each value $x \in \tilde{\mathcal{X}}$ the conditional mean value of Y, given $X=x$. This is often called a *mean value regression (function)*.

(b) Conditional quantiles. Specific regression functions have then the form $x \longrightarrow \mathrm{Q}_p(Y|X=x)$. The function assigns to each value $x \in \tilde{\mathcal{X}}$ the conditional p-quantile of Y, given $X=x$. It is, of course, possible to use several functions with different p-values simultaneously.

(c) Conditional frequencies. Specific regression functions have then the form $x \longrightarrow \mathrm{P}[Y|X=x](y)$. Following this approach one can define a specific regression function for each value $y \in \tilde{\mathcal{Y}}$. In particular, if Y is a binary variable with possible values 0 and 1, it suffices to consider a regression function $x \longrightarrow \mathrm{P}(Y=1\,|\,X=x)$.

8. Descriptive regression models

General as well as specific regression functions provide variants of statistical descriptions (derived from the distribution of a two-dimensional variable).

Table 1.3 Mean values, in specified income classes, of income and of expenditures on food (including beverages and tobacco) of private households in Germany 1998

		Expenditures on food	
Income class	*Mean income*	*in DM*	*in %*
under 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

Source: *Statistisches Jahrbuch für Deutschland* 2001: 573.

This corresponds to a statement made by V. R. McKim (1997: 7), "that regression is providing new facts, not interpretations or explanations of facts." Talk of regression models is more difficult to understand because quite different notions of such models exist. A basic distinction can be made between descriptive and analytical regression models. Descriptive regression models are used to provide simplified representations of general or specific regression functions, and their descriptive claim is derived from the regression functions they represent. In contrast, analytical regression models use the formal framework of regression functions to express hypotheses about substantial dependency relations. Since these models transcend the descriptive claims of regression functions derived from statistical variables, their formulation requires a different conceptual approach. In this text, analytical regression models will be considered as a variant of functional models; the discussion begins in Chapter 4.

A regression of private household's food expenditures on their income will be used to illustrate the notion of a descriptive regression model. The data, shown in Table 1.3, refer to private households who participated in a survey conducted in Germany in the year 1998. A variable $(X, Y) : \Omega \longrightarrow \tilde{\mathcal{X}} \times \tilde{\mathcal{Y}}$ can be used to provide a conceptual framework. Ω is the set of households participating in the survey; X records the net income, Y records the expenditures on food.[13] The data published by the *Statistisches Bundesamt* are mean values calculated for eight income classes, as shown in the table. Employing a further variable, Z, for the income classes ($\tilde{\mathcal{Z}} = \{1, \ldots, 8\}$), the table shows: $\bar{x}_j := \mathrm{M}(X|Z = j)$, the mean income in income class j, and $\bar{y}_j := \mathrm{M}(Y|Z = j)$, the mean value of expenditures on food in income class j. This then allows the consideration of a specific regression function

$$\bar{x}_j \longrightarrow \bar{q}_j := \frac{\bar{y}_j}{\bar{x}_j} \qquad (j = 1, \ldots, 8)$$

To each mean income value, the function assigns the mean value of the proportion of expenditures spent on food. This function is shown by the dots in Figure 1.2.

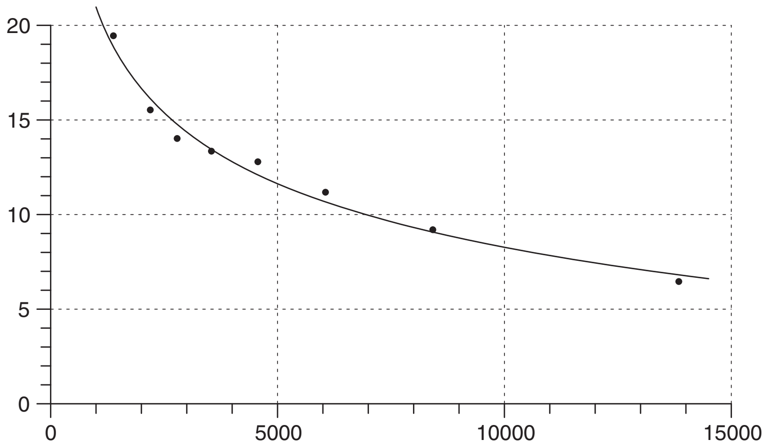


Figure 1.2 The dots correspond to the data in Table 1.3 (mean income values and proportion of income spend on food). The solid line shows the function $g(x; \hat{\alpha}, \hat{\beta}, \hat{\gamma})$ explained in § 8.

This is the original regression function as derived from the data. A descriptive model is a, in some sense simpler, representation of this function. In this example, one can consider a function $g : [0, 15\,000] \longrightarrow \mathbf{R}$, defined for some range of income values, that approximates the original regression function: $g(\bar{x}_j) \approx \bar{q}_j$. As suggested by the data, one might try model functions of the form

$$g(x; \alpha, \beta, \gamma) := \alpha\, x^{\beta} + \gamma$$

Using then the data from Table 1.3, one finds parameter values $\hat{\alpha} = 123.03$, $\hat{\beta} = -0.15$, and $\hat{\gamma} = -21.79$, that provide the best approximation (in a least squares sense). The solid line in Figure 1.2 shows the function with these parameter values.

Many different versions of descriptive regression models have been developed in the statistical literature. Often used are simple models for conditional mean values having the general form $\mathrm{M}(Y|X = x) \approx g(x; \theta)$; a well-known special case is the linear regression model $\mathrm{M}(Y|X = x) \approx \alpha + x\beta$. Obviously, also the first approach to the construction of regression functions, mentioned in § 7 (see (1.3)), leads to descriptive models.

9. Statistical and substantial conditions

Regression functions seem well suited to investigate how values of one variable depend on values of another variable. It is therefore important to understand that these functions cannot immediately provide information about dependency relations. The crucial distinction is between covariation and dependence. Regression functions provide descriptive information about covariation; in contrast, dependence is not a descriptive but a modal notion that refers to conditions for possibilities.

A conditional distribution like $\mathrm{P}[Y|X=x]$ neither requires nor implies that values of X are in any substantial sense a condition for values of Y. This follows already from the fact that both variables can be interchanged and no statistical criteria exist for a distinction between dependent and independent variables. Furthermore, the condition, $X=x$, just demarcates a part of the reference set for which the two-dimensional variable (X, Y) is defined. A quite different reasoning would be required in order to think of conditions in a substantial sense. One would need to consider the processes by which values of the dependent variable, Y, come into being; and only then it might become possible to think of values of X as referring to conditions on which these processes depend. (The argument will be continued in Section 2.1 dealing with statistical notions of structure.)

1.2 Relations

The statistical approach has been criticized as being "atomistic."[14] In a sense this is true. Beginning with statistical variables, the statistical approach can only consider properties which can be individually assigned to the members of a reference set. This is, however, only a formal requirement and does not exclude the possibility of considering relational properties. Given a reference set of people, say Ω, not only can each individual ω be characterized as being involved in relationships of some kind, also properties of the partners (e.g. their age) and of the relationships (e.g. their duration) can be recorded by formally attributing these properties to ω. However, the present section briefly introduces a simple extension of the conceptual framework that allows an explicit representation of relationships among the members of a reference set. This will also be helpful for the discussion, in Section 2.2, of an often assumed opposition between statistical and network approaches to social research.

1. Relational variables

As a general formal framework for the representation of relations between identifiable objects one can use relational variables having the form

$$R : \Omega \times \Omega \longrightarrow \tilde{\mathcal{R}}$$

Like statistical variables, relational variables are also functions. The domain consists of all pairs of objects which can be created from the members of Ω, a set of objects of some kind,[15] and the codomain $\tilde{\mathcal{R}}$ is a set of properties that can be used to characterize relationships between the members of Ω. In the most simple case is $\tilde{\mathcal{R}} = \{0, 1\}$, and the variable only records whether a relationship exists:

$$R(\omega, \omega') := \begin{cases} 1 & \text{if } \omega \text{ and } \omega' \text{ are connected (in a specified way)} \\ 0 & \text{otherwise} \end{cases}$$

Table 1.4 Participation of 18 women in 14 social events. The crosses indicate in which events a women participated

		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			

Source: Homans (1951: 83).

However, $\tilde{\mathcal{R}}$ can consist of any number of values referring to properties of possible relations between the members of Ω; and like statistical variables, also relational variables can be multidimensional consisting of two or more components.

2. *Construction of networks*

In this text, the word *network* will be used in an abstract meaning: something is a network if it can be represented by a relational variable. The reference set can consist of objects of any kind, often called the *nodes* of the network.

An example, often cited in the literature, will be used to illustrate the construction of networks. The data as published by G. C. Homans (1951: 83) are shown in Table 1.4. They provide information about the participation of 18 women in 14 social events.[16] In order to construct a network one first needs a reference set. In this example, the obvious choice is the set $\Omega := \{\omega_1, \ldots, \omega_{18}\}$ containing the names of the women. Networks can then be defined in different ways.

(a) For each of the events, $t = 1, \ldots, 14$, one can define a separate relational variable $R_t : \Omega \times \Omega \longrightarrow \{0, 1\}$ with $R_t(\omega, \omega') = 1$ if ω and ω' have both participated in the tth event.

(b) One can combine the information and define relations by the number of times two women have participated in the same event. The result is a single relational variable $R : \Omega \times \Omega \longrightarrow \{0, 1, 2, \ldots\}$ with $R(\omega, \omega')$ providing the number of events in which ω and ω' have both participated. Table 1.5 shows values of

Table 1.5 Adjacency matrix of a relational variable that records the number of times two women have participated in the same event. Calculated from Table 1.4

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

this relational variable in the form of an adjacency matrix $\mathbf{A} = (a_{ij})$ having coefficients defined by $a_{ij} := R(\omega_i, \omega_j)$.

These are just two possibilities to construct networks from the data in Table 1.4; one can easily invent further possibilities. The example thus shows that networks, similar to statistical distributions, are conceptual constructions even if they are derived from empirical facts.

3. *Formal descriptions of networks*

If a network has been explicitly defined by a relational variable, say $R : \Omega \times \Omega \longrightarrow \{0, 1, 2, \ldots\}$, it can be described in different ways. The main possibilities are as follows:[17]

(a) One can consider formal properties of the relationships, in particular, whether they are symmetrical, reflexive, and transitive.[18]
(b) One can investigate how often specified relationships occur and consider, for each $r \in \tilde{\mathcal{R}}$, a frequency $|\{(\omega, \omega') \,|\, R(\omega, \omega') = r\}| / |\Omega \times \Omega|$. If $\tilde{\mathcal{R}} = \{0, 1\}$ and $r = 1$, this is often called the *density* of the network. The general approach is obviously similar to the consideration of a statistical distribution.
(c) One can consider a great variety of formal properties of the network, for example the number of components and the existence of cliques and other types of subgroups.

(d) One can investigate how the individual nodes are embedded in the network and then characterize the nodes by accordingly defined properties (e.g. the node's number of connections to other nodes). The result is a statistical variable that can be described by a statistical distribution.

It is noteworthy that these are all *formal* descriptions of a network, independent of any substantial meaning of the relations referred to.

4. *Different kinds of relations*

Of course, understanding a network first of all requires an understanding of the relations used to construct the network. This is all the more necessary because these relations are often also conceptual constructions. The following distinctions point to some of the possibilities.

(a) *Comparative relations* between two or more objects are derived from properties which can be attributed to each object separately. For example, ω is older than ω'. In particular, comparative relations can be derived from a previously defined statistical variable, say $X : \Omega \longrightarrow \tilde{\mathcal{X}}$, by comparing $X(\omega)$ and $X(\omega')$.

(b) *Context-dependent relations* between two or more objects are derived from properties of a context to which the objects belong. These properties might refer to events or other kinds of facts. An *event-based relation* between two objects derives from an event in which both objects are involved in some way. For example, two persons talk together, or are both involved in a traffic accident, or have visited the same art exhibition, or have participated in an election in the same district. Event-based relations will be called *interactive relations*, or briefly *interactions*, if a communicative or in some sense causally relevant relationship is involved. Context-dependent relations can also be defined without referring to an event. For example, two villages are connected by a street, or two computers are connected by a cable that allows an exchange of data.

5. *Factual and modal views of relations*

An important distinction can be made between two views of relations. As an example, consider two computers connected by a cable that allows the exchange of data. On the one hand, this is an empirical fact that allows a factual statement; but the same fact, on the other hand, also refers to possibilities (the exchange of data) which *might*, or might not, become realized. Or think of another example: a person employed by a firm. Again this is an empirical fact that consists in the existence of an employment contract between the person and the firm; but again this fact also refers to modalities, in this example to certain kinds of behavior which *should* take place due to the contract. In the same way, when referring to a relation, one can distinguish between two views:

(a) A *factual view* that allows to state empirical facts constituting the relation (e.g. the existence of a cable conncecting two computers, or the existence of an employment contract); and
(b) a *modal view* that considers kinds of behaviors which, as a consequence of the relation, become possible, or probable, or normatively required.

It is noteworthy that the modalities referred to by a modal view of relations are different from empirical facts. The modal view must be distinguished, in particular, from a retrospective view of the history of a relation. For example, in a retrospective view one can consider whether, and to what extent, the connection between the two computers has in fact been used to exchange data, or one can describe the way in which the employee actually behaved (with respect to the contract that constituted the relation). In contrast, the modal view does not consider empirical facts (realized in the past), but modalities constituted by a relation, meaning behaviors that a relation makes possible, or probable, or normatively required.

2 Notions of structure

2.1 Statistical notions of structure
1. Structures as distributions
2. A version of social structure
3. Units of statistical structures
4. Data generating and substantial processes
5. Different kinds of macro facts
6. Statistical structures as conditions?

2.2 Taking relations into account
1. Relational notion of structure
2. Statistical reference sets
3. Statistical variables and relations
4. Dependence on structural conditions
5. Using induced relations

The previous chapter introduced the basic conceptual framework for statistical and relational descriptions. In the realm of social research those concepts can be used to provide descriptions of people and their living conditions. Further questions concern dependency relations: How are people dependent on social conditions? Arguments often employ notions of structure; for example:

> "Structure" is one of the most important and most elusive terms in the vocabulary of current social science. [. . .] The term structure empowers what it designates. Structure, in its nominative sense, always implies structure in its transitive verbal sense. Whatever aspect of social life we designate as structure is posited as "structuring" some other aspect of social existence – whether it is class that structures politics, gender that structures employment opportunities, rhetorical conventions that structure texts or utterances, or modes of production that structure social formations.
>
> (Sewell 1992: 1–2)

Leaving aside the obscure idea that structures can, in some sense, play an active role, it seems plausible that notions of social structure can sensibly be used to refer to those aspects of social conditions on which people effectively depend. But how can social structures be conceptually captured? An obvious idea would be to refer to the institutions that people have created as arrangements of their living conditions. The conceptual approach of both statistical and relational notions of structure is, however, quite different. These notions derive from statistical and relational variables and therefore immediately give rise to the question in which sense, if at all, they can provide an understanding of social conditions on which people depend.

The first section of the present chapter considers statistical notions of structure. It is argued that these versions of structure are essentially conceptual constructions; and this argument then leads to the question whether, or in which sense, statistical structures can be understood as conditions on which people (or more abstractly defined social processes) depend. The second section argues that there is no clearcut opposition between statistical and relational notions of structure, and several possibilities exist for the statistical approach to take into account relational properties.

2.1 Statistical notions of structure

1. Structures as distributions

The word 'structure' does not have a unique meaning. In statistical social research it is most often used synonymously with 'statistical distribution.' In this sense, for example, the age structure of a population simply means the distribution of a statistical variable that records for each member of the population its age. Correspondingly, social structure in its statistical sense often means the distribution of statistical variables that record relevant properties, like education and income, or an assignment of categories from a previously constructed stratification.

It is noteworthy that some of the ideas often connected with an understanding of "structure" do not fit with its statistical notion. Consider the following remark made by G. C. Homans:

> [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.
>
> (Homans 1976: 54)

In contrast to this understanding, the statistical notion of structure has no relational connotations at all. A second difference concerns the idea that 'structure' often means a set of relatively permanent conditions (for processes of some kind). Not only has the statistical notion of structure no implications with respect to temporal permanence. Most often it is also not possible to think of a substantial process for which a statistical structure (distribution) is in some sense a condition. (This argument will be continued in § 6.)

2. A version of social structure

The macrostructural approach proposed by Peter M. Blau provides a good example of a statistical notion of social structure. The following statement explains the general idea:

> 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 1994: 1)

The dimensions can refer to any properties by which the members of a society can be distinguished. Blau often speaks of "social positions," but actually means "any difference among people in terms of which they make social distinctions among themselves in social intercourse" (Blau 1994: 3). Important is the interest in relations between the various dimensions because this requires not to merge the different dimensions into a one-dimensional classification. Correspondingly, Blau proposes the following definition:

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

As a formal framework one can use a multidimensional statistical variable

$$(X_1, \ldots, X_m) : \Omega \longrightarrow \tilde{\mathcal{X}}_1 \times \cdots \times \tilde{\mathcal{X}}_m \tag{2.1}$$

The reference set, Ω, represents the population, and the property spaces, $\tilde{\mathcal{X}}_1, \ldots, \tilde{\mathcal{X}}_m$, refer to the dimensions to be used for distinctions between the members of Ω. What Blau calls a social structure is formally identical with the statistical distribution of $(X_1, \ldots, X_m)$. Relations between the various dimensions can then be studied with measures of correlation (Blau 1994: 5) or, more generally, with various kinds of regression functions.

3. Units of statistical structures

Blau's notion of social structure refers to populations consisting of human individuals. It is obvious, however, that a completely analogous approach can be used to define statistical structures for other kinds of units, e.g. households, firms, or regions. Units are simply the individual members of the reference set used to define the statistical variable (2.1).

One should be aware, however, of a certain ambiguity that, for example, occurs in the following passage (Blau 1974: 615–6):

> 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. [...] 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.

At first sight this seems to contradict the understanding that Blau's notion of social structure refers to a population (a reference set consisting of human individuals). The contradiction dissolves, however, if one explicitly distinguishes between units and positions (in the meaning supposed by Blau). The formal framework of a statistical variable, say $X : \Omega \longrightarrow \tilde{\mathcal{X}}$, supports the distinction. The elements of the reference set, Ω, are the units, and the elements of the (possibly multidimensional) property space, $\tilde{\mathcal{X}}$, are the positions. The variable, X, also induces a partition of the reference set: To each position $x \in \tilde{\mathcal{X}}$ corresponds a subset

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

consisting of units sharing the position x. Although these subsets are conceptually different from the positions (as indicated by Blau), the frequency of the positions obviously equals the size of the subsets.

A real source of obscurity is, however, Blau's statement that his notion of social structure refers to "component parts *and their relations*" (emphasis added). As already remarked in § 1, this relational rhetoric does not fit with a statistical notion of structure. While the fact that two or more individuals share the same position (in Blau's sense) might be used to construct comparative relations, these relations only exist on a conceptual level and should not be confused with relations in any substantial sense (see Section 1.2, § 4).[1] On the other hand, while Blau's talk of "social interaction and communication" clearly refers to substantial relations, *these* relations do not play any role in the statistical notion of social structure.

4. Data generating and substantial processes

In order to understand better statistical notions of structure, one might ask: how do statistical facts come into being? Consider the following picture:

○ ○ ● ○ ● ● ○ ○ ○ ●

What the picture shows (10 objects of which 4 are black and 6 are not black) is obviously not already a statistical fact. A statistical fact only comes into being as the result of a specific procedure which consists of three steps: (a) conceptualization of a statistical variable; (b) a (real or fictitious) *data generating process* that provides values of the variable; and (c) calculations which finally create the statistical fact (the variable's distribution or some quantities, or functions, derived thereof).

In the example, the three steps can easily be performed. Using a reference set $\Omega := \{\omega_1, \ldots, \omega_{10}\}$ to represent the objects, and a property space, say $\tilde{\mathcal{Y}} := \{0, 1\}$, to distinguish black (1) and non-black (0) objects, one can define a statistical variable $Y : \Omega \longrightarrow \tilde{\mathcal{Y}}$ that assigns to each object a specific value in the property space $\tilde{\mathcal{Y}}$. Then follows the second step providing the data. In this example this is easily done, and the result can be made available in the following form:

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

This table provides the data for the third step, that is, the calculation of the distribution P[Y]. In this example no extensive calculations are necessary and one directly gets the result: $\mathrm{P}(Y = 0) = 0.6$ and $\mathrm{P}(Y = 1) = 0.4$.

These statements formulate the statistical fact; and the example thus provides one possible answer to the initial question: Statistical facts come into being as the result of a theoretical and practical procedure that consists of the three mentioned steps. The example also shows that a data generating process presupposes that the facts to which the data refer already exist. These facts, which correspond to the individual elements of the reference set, will subsequently be called *micro facts*. Obviously, they are different from the derived statistical facts, and it is necessary, therefore, to distinguish two questions:

(a) How did the micro facts (supposed to exist in the empirical world) come into being? (In the example: How did the 10 black and non-black objects come into being?)
(b) How do statistical facts (referring to already existing micro facts) come into being?

The following diagram illustrates the distinction:

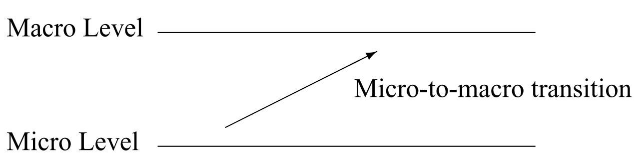
$$\left.\begin{array}{l} -\!\!\rightarrow\; Y(\omega_1) \\ \qquad\;\;\vdots \\ -\!\!\rightarrow\; Y(\omega_n) \end{array}\right\} \Longrightarrow \mathrm{P}[Y]$$

Arrows of the form $--\rightarrow$ are intended to hint at the *substantial processes* that create the micro facts referred to by $Y(\omega_1), \ldots, Y(\omega_n)$. Given that these micro facts already exist, a quite different process, indicated by the arrow $\Longrightarrow$, creates the statistical fact $\mathrm{P}[Y]$. Thus, three processes should be distinguished:

(1) *Substantial processes* generating the micro facts assumed to exist correspondingly to the individual elements of a reference set;
(2) *data generating processes* generating information (data) about the micro facts;[2] and
(3) *calculations* which, beginning with the results of a data generating process, create statistical facts.

5. *Different kinds of macro facts*

The foregoing talk of micro facts might suggest calling statistical facts correspondingly macro facts. However, it is important, then, to distinguish statistical facts from other kinds of macro facts which can be considered as being empirical facts because they result from the interaction of people. A diagram proposed by J. S. Coleman (1990: 8–23) can serve to illustrate the difference:



Following Coleman, micro-level propositions refer to the behavior of (individual or corporate) actors, and macro-level propositions refer to facts which, in some sense, result from the actions performed at the micro level. But given this understanding, statistical facts do not correspond to the macro level. The reason is not that the micro facts from which statistical facts are derived most often do not consist in individual actions; the interpretation of the micro level may be extended to allow a reference to any kind of micro facts. The crucial point is that statistical facts do not result from processes which can be theoretically linked to the behavior of micro-level actors; they rather result from observations and calculations which presuppose that the substantial processes have already taken place.[3]

6. *Statistical structures as conditions?*

Can statistical structures be used to capture conditions on which the behavior of people effectively depends? Dealing with this question, one should be aware of a possible confusion. Consider, as a simple example, a variable $X : \Omega \longrightarrow \{0, 1\}$. Ω is a set of people, and X records whether members of Ω live in a household that has available a washing-machine (1) or otherwise (0). For each $\omega \in \Omega$, $X(\omega)$ obviously documents an aspect of ω's living conditions. However, these are values of a statistical variable, not structures. The statistical structure, in this example, refers to the distribution $P[X]$, and this distribution is not a condition on which any member of Ω effectively depends.

The question remains whether individuals sometimes also depend on *structural conditions*, that is, conditions defined in terms of statistical distributions. Using again a variable $X : \Omega \longrightarrow \tilde{\mathcal{X}}$ as a formal framework, the question is whether there are examples in which one can think of the distribution $P[X]$ as being a relevant condition for …? Three cases can be distinguished.

(a) One can think of an individual actor, say A, and imagine that A's behavior in some way depends on the distribution $P[X]$. Although it is not required that A is a member of Ω, it must be assumed that $P[X]$ refers to a situation which contains A and is relevant for A's behavior. As an example, one can think that A is teacher of a class, represented by Ω, and $P[X]$ describes the proportions of boys and girls in the class.

(b) The example directly leads to a second possibility. It is obviously possible to say that $P[X]$ is also a condition for the individual members of Ω. Of course, the way in which $P[X]$ is an effective condition for each individual $\omega \in \Omega$ may also depend on other circumstances, including ω's properties (e.g. ω's sex).

(c) A third case is different in that $P[X]$ no longer characterizes situations in which individual actors actually exist. As a modification of the previous example, assume that Ω consists of two classes, Ω_1, consisting only of boys, and Ω_2, consisting only of girls. $P[X]$, the sex distribution in the reference set $\Omega = \Omega_1 \cup \Omega_2$, no longer, then, refers to effective conditions of the individual members in the two classes.

The conclusion is that statistical distributions can sometimes be used to define structural conditions; a crucial precondition is, however, that these conditions can be linked to situations in which individual actors (or other kinds of objects) actually exist.

2.2 Taking relations into account

1. Relational notion of structure

The following quotation explains a *relational notion of structure*: "A structure is a configuration of parts, and a structural description is a characterization of the way the components in a set are interrelated." (Hernes 1976: 518) The notion is obviously very general and abstract. It might refer to a *structured unit*, meaning a unit that can be considered as consisting of identifiable parts which are in some way related. For example, in social research, these are couples, households, firms, and other kinds of organizations. But the notion can also be used to refer to any kind of network (as defined in § 2 of Section 1.2). To describe a relational structure then means to describe the network.

2. Statistical reference sets

Reference sets of statistical variables are most often considered as consisting of unrelated units.[4] In contrast, the network approach in social research starts from the consideration of systems consisting of elements which are in some way related. These systems are formally represented as networks which then become the objects of description and analysis. It therefore seems that there is a fundamental difference between a statistical and a network approach to social research. However, there are several reasons why the contradistinction is not clearcut.

First, statistical variables can also be used to characterize the elements of a statistical reference set by relational properties. This is obvious if these elements are already defined as structured units (e.g. households or firms). But even if not, they can be characterized by properties of structured units to which they belong. For example, one can characterize individual persons by the size of the household to which they belong. This obviously does not require that all household members have an explicit representation in the statistical reference set.

3. Statistical variables and relations

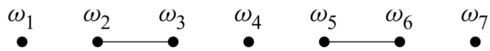
A second consideration concerns the simplistic opposition of statistical and relational notions of structure. Actually, given a statistical variable $X : \Omega \longrightarrow \tilde{\mathcal{X}}$, one can always derive at least one relational structure for the reference set Ω. The most simple one would be the equivalence relation induced by X, that is, a partition of Ω into classes of elements having identical values of X. This is the basic statistical relation: Two units, ω' and ω'', are *statistically equivalent* (w.r.t. X) if $X(\omega') = X(\omega'')$.

It should be added that the relevant meaning of this notion derives from a presupposition: that in defining the property space of a statistical variable one does not use proper names which identify the individual units. In other words, a reference set is considered as a set of generic units, only distinguishable by values of explicitly defined variables. While there are sometimes exceptions in small-N research, I take this presupposition as fundamental to the statistical approach to social research.[5]

On the other hand, assume that a relational structure for Ω is explicitly given by a relational variable $R : \Omega \times \Omega \longrightarrow \tilde{\mathcal{R}}$. Instead of statistically equivalent units one can then identify units that have an equivalent position in the relational structure. We briefly consider two definitions.

(a) Two units, ω' and ω'', are *structurally equivalent* (w.r.t. R) *iff* $R(\omega', \omega) = R(\omega'', \omega)$ and $R(\omega, \omega') = R(\omega, \omega'')$ for all $\omega \in \Omega \setminus \{\omega', \omega''\}$.[6]
(b) Two units, ω' and ω'', are *automorphically equivalent* (w.r.t. R) *iff* there exists an automorphism $\pi : \Omega \longrightarrow \Omega$ with $\omega'' = \pi(\omega')$.[7]

An important difference concerns the question whether individual units can be identified (by using their labels). As an illustration consider a reference set consisting of seven units with a relational structure as follows:



There are two classes of automorphically equivalent units: $\{\omega_1, \omega_4, \omega_7\}$ and $\{\omega_2, \omega_3, \omega_5, \omega_6\}$. Using instead the notion of structural equivalence would allow to further distinguish between $\{\omega_2, \omega_3\}$ and $\{\omega_5, \omega_6\}$. However, the statement that, say, ω_2 and ω_5 are not structurally equivalent obviously requires that one can distinguish between ω_3 and ω_6, and this is only possible by using their labels.

One should therefore prefer the notion of automorphic equivalence when following a statistical view that assumes that individual units cannot be identified (but only characterized by values of variables). The statistically relevant information provided by the relational structure R can then be represented by a statistical variable that assigns each element of Ω to an equivalence class of the automorphic equivalence relation. In the example, the statistical variable would distinguish between $\{\omega_1, \omega_4, \omega_7\}$ and $\{\omega_2, \omega_3, \omega_5, \omega_6\}$; and this partition would provide a complete formal characterization of how the units are part of the relational structure: Belonging to the first class a unit is isolated, and belonging to the second class a unit is connected with another one. Of course, one normally has additional information about the relation (e.g. being married), and this then provides additional meaning for the equivalence classes.

The consideration shows that the information which a relational structure contains about the individual members of the reference set might well be captured by statistical variables. This undermines any simple opposition between a statistical and a relational (network) approach. Of course, the leading question might concern properties of a system, represented by the relational structure, as a whole. A statistical approach that views the elements of the system as individual elements of a statistical reference set is then clearly not appropriate, and proper methods of network analysis should be applied. But our present question is different: Starting from the idea that the individual elements of a statistical reference set might depend on structural relations, how can this be taken into account? The consideration has shown that the conceptual framework of statistical variables might well suffice.[8]

There are, of course, limitations. In accordance with the definition given in § 1, a reference set Ω can be called a *structured unit* if there is a relational structure and there are no (or almost no) automorphically equivalent elements. A statistical variable that records membership in equivalence classes will then assign all (or almost all) elements of Ω a unique value, and in this sense allows the identification of them. Such a variable cannot be used for statistical analyses because the notion of a frequency distribution (although formally still applicable) loses its sense. The idea that statistical variables can be used to capture positions in a relational structure therefore has limitations.

4. Dependence on structural conditions

A further consideration takes up the idea that individual units might depend on properties of a relational structure. This is often seen as one of the basic and most important ideas of a network approach.[9] However, since the notion of a relational structure is extremely general and abstract, one will need an explicit argument for the assumption that a relational structure constitutes effective conditions for its elements.

Whether, and how, this assumption can be justified depends on the application context. But remembering the discussion in Section 1.2, there is at least the general requirement that relations allow a modal interpretation. Therefore, comparative relations can most often not be interpreted as effective conditions. Instead one has to consider context-dependent relations that characterize environments of the members of the reference set Ω. For each individual unit ω, the specification of the relevant environment might require a reference to other members of Ω; but this is obviously of no importance for the question whether or not the specified environment constitutes an effective condition.

The argument parallels the one made at the end of Section 2.1. In any case, using statistical or relational notions of structures, it is important to refer to local environments of the individual units in order to specify conditions on which these units effectively depend.

5. Using induced relations

In statistical social research one rarely has data about context-dependent relations among the individual units of a reference set. It might well be possible, however, to incorporate information derived from relations defined on property spaces. Assume that a relational structure $R : \tilde{\mathcal{X}} \times \tilde{\mathcal{X}} \longrightarrow \tilde{\mathcal{R}}$ is defined for the property space of a statistical variable $X : \Omega \longrightarrow \tilde{\mathcal{X}}$. This induces a relation on Ω, say $R^* : \Omega \times \Omega \longrightarrow \tilde{\mathcal{R}}$, defined by $R^*(\omega', \omega'') := R(X(\omega'), X(\omega''))$.[10] For example, $\tilde{\mathcal{X}}$ may comprise a set of spatial locations, and R provide information about their spatial distances. $R^*(\omega', \omega'')$ is then a measure of the spatial proximity of ω' and ω''. (An application will be discussed in Section 8.2 when dealing with diffusion models.)

3 Processes and process frames

3.1 Historical and repeatable processes
1. Some notions of process
2. Historical processes
3. Process frames
4. Processes and rules

3.2 Time series and statistical processes
1. Time axes
2. Timed process frames
3. Statistical processes
4. Aggregation of individual processes

3.3 Stochastic process frames
1. Random generators
2. Aleatoric probability
3. Stochastic variables
4. Stochastic process frames

Having introduced elementary statistical and relational concepts in Chapter 1 and discussed corresponding notions of structure in Chapter 2, this chapter deals with some possibilities to conceptualize processes. The first section is concerned with a distinction between historical processes and process frames that are used to define repeatable processes. The second section deals with schematic process frames that are based on explicitly defined time axes, in particular time series and statistical processes. The third section explains a notion of stochastic process frames that will be used in later chapters for the discussion of stochastic models.

3.1 Historical and repeatable processes

1. Some notions of process

In its most general sense, ‘process’ only means that something develops in time. The word can therefore be used in many different more specific meanings. Here are some possibilities:

- The word can refer to the development of states (properties) of identifiable objects; for example the development of the employment status of a specified person during a specified time span.
- The word can refer to the development of states attributable to a plurality of objects. Using statistical concepts, this leads to statistical processes in the sense of temporal sequences of statistical variables.
- Different notions of process emerge if one begins, not with states, but with events. In a general sense, the notion of process then means a temporally structured series of events.
- Depending on the kinds of events more specific notions can be used. In particular, one can think of behavioral processes which consist of temporally structured activities of (one or several) individuals.

An obvious question always concerns how to delineate the events, or objects and their properties, that are part of the process. Observations will not provide an answer because the possibilities of considering objects and events as being part of a process are virtually unlimited. It is therefore proposed to understand 'process', not as an ontological category (something that exists in addition to objects and events), but as belonging to the ideas and imaginations of humans aiming at an understanding of the occurrences they are observing. Put somewhat differently, it is suggested to understand processes as conceptual constructions. This is not to deny that processes can meaningfully be linked to observations of objects and events; but this will then be an indirect link: one can observe objects and events, but not processes (even if they are constructed from observations).

2. Historical processes

A process will be called *historical* if its construction refers to an empirically identifiable development in the empirical world. The notion presupposes a human practice that provides the point of view from which a historical process can be identified. Correspondingly, the notion also presupposes an understanding of time that is grounded in human practice, in particular, the distinction between temporal modalities: a past that consists of realized facts which cannot be changed, a future that consists of possibilities which might become realized, and a transitory present in which some of the possibilities become realized and thereby become part of the past.[1]

As proposed here, the construction of historical processes does not require any specific conceptual tools. One can construct historical processes as behavioral processes or, more generally, as a temporally structured series of events. But also statistical concepts can be used to construct, for example, a historical process that refers to the development of the population in Germany during a specified period.

Since it is required that historical processes can be empirically identified, these processes always have a beginning and an end in the historical time. It follows that a historical process always refers to the past of the human practice by which the process is identified. As examples one can think of a conversation, a walk, the formation of a building, or the population development during a specified period. Obviously, there is no limit to imagining different historical processes.

As the examples also show, there is no requirement that historical processes occupy a "long" time span. The distinction between processes and events that will be made in this text does not result from different temporal extensions but rather concerns the ontological status. Events are conceived as occurrences that can be observed in the empirical world; in their ontological status they are therefore similar to empirically identifiable objects. Processes, on the other hand, are conceptual constructions and do not have an empirical existence on their own.[2]

3. Process frames

A *repeatable process* is a process that can take place several times in a similar way. For example, walking from home to work. In this understanding, the notion only requires that one can think of several processes which are in some way comparable.

The notion of repeatable processes suggests an important distinction between processes and process frames. While a process is a unique development, a *process frame* provides a conceptual framework that allows one to think of repeatable processes as realizations of the process frame. These realizations will be historical processes if they actually take place in the empirical world; it is quite possible, however, also to imagine a fictitious process as being the realization of a process frame. The relevant distinction is therefore not between historical and repeatable processes, but between historical and fictitious processes. If a historical (or fictitious) process is termed repeatable, the meaning is that the process shall be viewed as the realization of a (previously constructed) process frame.

- As examples one can think of computer programs (or more abstract, algorithms). One obviously has to distinguish between a computer program, in the sense of a process frame, and the repeatable processes that will take place when the program gets started. Each of these processes will be a historical process that takes place in a specific spatial and temporal context.[3]
- Also games, like chess, provide easily understandable examples. On the one hand there is a process frame defined by the rules of the game, on the other hand there are the realized games which can be imagined as fictitious processes or actually performed as historical processes.
- One can also think of behavioral processes which are regulated to some extent, for example, the preparation of dishes or the provision of a medical diagnosis. The regulation (consisting of rules) provides a process frame to be distinguished from the behavioral processes that take place as real or imaginary realizations.

- Finally, human life courses can also be considered as repeatable processes. Of course, no one can repeat his or her life; but the idea simply is to view life courses (of a specified set of individuals) as being realizations of a process frame. The process frame might be a pure theoretical construction or correspond to institutionalized regulations.

The examples show that repeatable processes can be quite different. While computer programs are examples of mechanical processes, games are behavioral processes which can be influenced by human actors.[4] The examples also show that the scope of potential realizations which a process frame allows can be quite different. While algorithms that uniquely determine a process provide an extreme case, most process frames allow a more or less broad spectrum of different realizations.

4. Processes and rules

Process frames can often be described by rules. In these cases one might say that the realized processes are "governed by rules." The formulation is, however, ambiguous.

First of all, that a process is governed by rules does not imply that the process is determined by rules. Gilbert Ryle (1949: 77–9) has shown this by referring to games of chess: although governed by rules, the rules do not determine specific courses of the game. The example also shows that rules cannot be understood as causes which, in some sense, guarantee that a process develops in accordance with the rules; the rules which exist for games of chess cannot guarantee that players observe the rules.

This is quite independent of the scope of possible processes allowed by a set of rules and is true also for mechanical processes. For example, also a computer program cannot guarantee that its activation generates a process that conforms to the program. The actual development of the historical process which follows the program's activation rather depends on many conditions not taken into account by the program (e.g. the physical conditions of the hardware). For an understanding of the idea that processes are governed by rules it is, nevertheless, of importance whether actors only initialize a process or also influence its development.

In both cases rules can be part of a model, that is, a process frame which is constructed from the point of view of an observer for the consideration of possible or actually realized processes. Rules of this kind (called *modeling rules* in this text) govern the processes which are possible in the framework of the model. However, if processes involve actors it becomes possible and is often the case that these actors also have knowledge about rules. These are not modeling rules which govern the possible realizations of a model but rules from the point of view of actors. Nevertheless, the development of a process also depends on these *actor's rules* (as they will briefly be called), or more precisely, it depends on how the actors observe, or not observe, the rules in their activities.

3.2 Time series and statistical processes

Processes and process frames can be constructed in several different ways. This section considers time series and statistical processes which are based on explicitly defined time axes.

1. Time axes

A time axis is a formal representation of time. Mainly two versions can be distinguished. One approach treats time as a sequence of *temporal locations* (e.g. seconds, hours, days, months, or years) and represents a time axis by integral numbers with an arbitrarily fixed origin. This is called a *discrete time axis*. Another approach treats time as a continuum (a "continuous flow of time") and represents a time axis by the set of real numbers. This is called a *continuous time axis*.

In this text a discrete time axis will be used. This has the advantage that one can think of a *sequence* of temporal locations. The notation is $\mathcal{T} := \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$, and it is assumed that the integers represent temporal locations (suitably specified in the application context). The notation $\mathcal{T}^*$ is used if only a part of $\mathcal{T}$ is needed; it is always a contiguous and (if not explicitly said otherwise) also a finite subset of $\mathcal{T}$.

2. Timed process frames

Having available the notion of a time axis one can immediately construct timed process frames. The following definition will be used: Given a time axis $\mathcal{T}^*$, a *timed process frame* is a corresponding sequence of logical variables (of some kind).[5] Realizations will accordingly be called *timed processes* (or briefly processes). Mainly two types of timed process frames will be distinguished.

- *Process frames for time series.* These process frames can formally be represented by sets $\{\ddot{X}_t \,|\, t \in \mathcal{T}^*\}$ where $\ddot{X}_t$ are logical variables for real numbers or vectors. Realizations are called accordingly *simple* or *vectorial time series*, formally $\{x_t \,|\, t \in \mathcal{T}^*\}$. Note the distinction between the logical variables, $\ddot{X}_t$, used to represent the process frame, and its possible values denoted by x_t. The lower-case symbols x_t are used not as logical variables, but as names representing specific real numbers or vectors.
- *Process frames for statistical processes.* These process frames can formally be represented by sets $\{X_t^* \,|\, t \in \mathcal{T}^*\}$ where X_t^* are logical variables for statistical variables (i.e. variables having statistical variables as possible values).[6] Realizations of these process frames are sequences of statistical variables and will be called *statistical processes*.

3. Statistical processes

As the word is used in this text, a statistical process is a sequence of statistical variables, X_t, with t referring to a time axis $\mathcal{T}^*$.[7] It will be assumed that there is a common property space, $\tilde{\mathcal{X}}$, for all variables X_t. Three types of statistical processes can then be distinguished. (a) If the reference sets of the variables can be different in each temporal location, formally

$$X_t : \Omega_t \longrightarrow \tilde{\mathcal{X}} \tag{3.1}$$

the process is called *synchronously aggregated*. (b) If there is a single reference set, Ω, that does not change while the process continues, formally

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

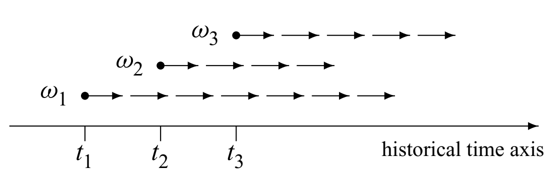
the process is called *temporally aggregated*. (c) A further possibility is that the reference sets can change while the process continues but, different from case (a), one can identify individual members across temporal locations.

In case (a), $\mathcal{T}^*$ is most often part of a historical time axis and the statistical process can then be considered as a historical process that develops during a specified historical period. In case (b), the time axis often does not correspond to a specific historical time span but is a theoretical construct that is used to define the statistical process; and most often the time axis is then formally specified as $\mathcal{T}_0 := \{0, 1, 2, \ldots\}$.

4. *Aggregation of individual processes*

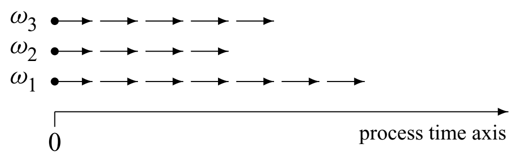
Temporally aggregated statistical processes can be viewed as resulting from an aggregation of individual processes. As is obvious from (3.2), the statistical process equals a collection of individual processes, $\{X_t(\omega) \,|\, t \in \mathcal{T}^*\}$.

While in this expression the individual processes are already defined on a common process time axis, $\mathcal{T}^*$, it is quite possible that they derive from individual processes that begin in different temporal locations on a historical time axis; to illustrate:



As an example, one can think of marriage episodes; the diagram then shows three marriage episodes which begin in the temporal locations t_1, t_2, and t_3, respectively.

The diagram also shows that individual processes can be aggregated in different ways. (a) One can define Ω as consisting of individuals whose individual processes began in the same temporal location on the historical time axis; for example: individuals who married in the same year. (b) One can define Ω as consisting of individuals whose individual processes ended in the same temporal location on the historical time axis; for example: individuals who became divorced in the same year. (c) One can define Ω as consisting of individuals whose individual processes began (or ended) during a longer historical period. Possibility (a) corresponds to a *cohort approach*, based on the understanding that "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." (Glenn 1977: 8) In any case, if one is interested in statistical statements about a collection of individual processes, they must be aggregated on a common process time axis. In the example:



it is assumed that all individual processes begin in the same temporal location 0 of a time axis $\mathcal{T}_0 := \{0, 1, 2, 3, \ldots\}$; and this allows for the consideration of the individual processes as if they develop in parallel.

3.3 Stochastic process frames

If the goal is to describe time series, or statistical processes, as historical processes for which data are available, it is not necessary to conceive of the processes as realizations of suitably constructed process frames. Explicitly constructed process frames are only required for the construction of (dynamic) models serving to consider repeatable processes and how they might be governed by rules. An extensive discussion of such models begins in Chapter 4. Since these models often use stochastic process frames the present section briefly introduces the basic notions.

1. Random generators

For the definition of stochastic process frames a specific notion of probability is used that is called aleatoric probability and refers to *random generators*. The following diagram can serve as an explanation.[8]

$$\text{Activation} \longrightarrow \boxed{\text{Random generator}} \longrightarrow \text{Outcome}$$

- First of all, a random generator is a method to generate outcomes. This implies that a random generator is not by itself an actor, but presupposes an actor who, by using the random generator, generates outcomes. Think for example of a die. In order to generate an outcome somebody must throw the die.
- The description of a random generator therefore consists in the description of a procedure to be followed in the generation of outcomes. This might include a description of specific devices (a die, for example) and how they shall be used.

- The notion of a random generator implies that it can be used any number of times to generate outcomes.
- A random generator has two or more possible outcomes; whenever activated, the random generator generates exactly one of its possible outcomes. The specification of the set of possible outcomes is part of the definition of the random generator.
- It is required that the process that begins with the activation of a random generator and creates the outcome develops independently of all historical circumstances, including, in particular, the time, the place, and the actor who initiates the process but cannot influence its further development.[9] Of course, this is to be understood as an ideal requirement which can be satisfied only approximately by real random generators.
- The requirement just mentioned also includes that processes which begin with the activation of a random generator develop independently of all processes and their outcomes which have been realized earlier (with the same or with any other random generator). Again, it is an ideal requirement: an ideal random generator has no memory.

2. Aleatoric probability

The notion of aleatoric probability serves to describe random generators. Referring to a random generator $\mathcal{G}$, the formal definition proceeds in two steps.

In a first step one defines a set, say $\tilde{\mathcal{Z}}$, that represents the possible outcomes. Analogously to the property spaces of statistical variables, it will be assumed that one uses a numerical representation and $\tilde{\mathcal{Z}}$ can be treated as a subset of the real numbers. For example, $\tilde{\mathcal{Z}} = \{1,\ldots,6\}$ if $\mathcal{G}$ refers to the generation of numbers with a single die. In a second step one specifies a probability measure, that is, a function

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

which associates with each subset $\tilde{Z} \subseteq \tilde{\mathcal{Z}}$ (each element of the power set $\mathcal{P}(\tilde{\mathcal{Z}})$) a number $\Pr[\mathcal{G}](\tilde{Z})$, to be interpreted as the probability that the activation of $\mathcal{G}$ generates an outcome in $\tilde{Z}$, and which satisfies the following requirements:

(a) for all $\tilde{Z} \subseteq \tilde{\mathcal{Z}}$: $\quad 0 \leq \Pr[\mathcal{G}](\tilde{Z}) \leq 1$

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

(c) for all $\tilde{Z}, \tilde{Z}' \subseteq \tilde{\mathcal{Z}}$: if $\tilde{Z} \cap \tilde{Z}' = \emptyset$ then:
$\Pr[\mathcal{G}](\tilde{Z} \cup \tilde{Z}') = \Pr[\mathcal{G}](\tilde{Z}) + \Pr[\mathcal{G}](\tilde{Z}')$

These requirements formally equal those for a statistical distribution, and probability measures are therefore formally equivalent with frequency functions. Unfortunately, this fact easily obscures the fundamental difference in meaning: While a statistical distribution refers to a collection of facts which have been realized in the past, a probability measure refers to possibilities which, in the case of aleatoric probability, might become realized by using a random generator. This text therefore uses different symbols, $\mathrm{P}[X]$ for the distribution of a statistical variable X, and $\mathrm{Pr}[\mathcal{G}]$ for the probability measure of a random generator $\mathcal{G}$.

A further question concerns the numerical specification of probabilities. However, when aleatoric probabilities refer to random generators which are explicitly known there is no estimation problem; numerical values can directly be derived from the known properties of the random generator.

3. Stochastic variables

To allow flexible references to random generators it is practical to use stochastic variables (also called random variables). A *stochastic variable* is a function

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

that has as its domain, $\tilde{\mathcal{Z}}$, the set of possible outcomes of a random generator; the codomain, also called the *range* of $\dot{X}$, is any property space (which can be identical with $\tilde{\mathcal{Z}}$). If then a stochastic variable $\dot{X}$ is defined by referring to a random generator $\mathcal{G}$, one can associate with each subset $\tilde{X} \subseteq \tilde{\mathcal{X}}$ an induced probability

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

The function $\mathrm{Pr}[\dot{X}]$ is called the *probability distribution*, or *probability measure*, of $\dot{X}$. Accordingly, one says that $\mathrm{Pr}[\dot{X}](\tilde{X})$ is the probability that $\dot{X}$ realizes a value in the set $\tilde{X}$. Often the following equivalent notations are used:

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

As already mentioned, from a purely formal view there are no differences between probability distributions of stochastic variables and statistical distributions of statistical variables. This text, nevertheless, uses different notations ($\dot{X}$ for stochastic variables, and X for statistical variables) to remind of quite different meanings:

- A statistical variable $X : \Omega \longrightarrow \tilde{\mathcal{X}}$ has as its domain a set Ω which consists of real or fictitious objects or situations. In contrast, the domain of a stochastic variable $\dot{X}$ consists of the set of possible outcomes of a random generator.
- Correspondingly, a statistical variable records properties of objects or situations which are actually realized or assumed to be realized. A stochastic variable, in contrast, refers to processes by which outcomes might come into being. Using stochastic variables therefore presupposes a modal consideration.
- This implies that the distributions also have a different meaning. While a statistical distribution $\mathrm{P}[X]$ represents a (statistically constructed) fact derived from realized properties of the members of a reference set, the probability distribution $\mathrm{Pr}[\dot{X}]$ describes a random generator, that is, a method that can be used to generate facts.

4. Stochastic process frames

Using stochastic variables one can construct stochastic process frames in many different ways. Here two timed versions will be considered which are defined analogously to process frames for time series (see Section 3.2, § 2). Both are based on a process time axis $\mathcal{T}^* = \{0, 1, 2, 3, \ldots\}$. The first version is simply defined as a sequence of stochastic variables:

$$\{\dot{X}_t \mid t \in \mathcal{T}^*\} \tag{3.1}$$

The second version is defined as a sequence

$$\{(\dot{X}_t, \ddot{Z}_t) \mid t \in \mathcal{T}^*\} \tag{3.2}$$

In each temporal location, t, there is a stochastic variable $\dot{X}_t$ and also a non-stochastic variable $\ddot{Z}_t$.

Both versions establish frames for processes that emerge if the variables $\dot{X}_t$, or $(\dot{X}_t, \ddot{Z}_t)$, get specific values. The results are time series

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

Obviously, a stochastic process frame allows one to think of, or actually generate, any number of such time series. Referring to n realizations of a process frame $\{\dot{X}_t \mid t \in \mathcal{T}^*\}$, they can also be viewed as the elements of a temporally aggregated statistical process $X_t : \Omega \longrightarrow \tilde{\mathcal{X}}$. The elements of Ω identify the n time series, and $X_t(\omega)$ is the value of the time series identified by ω in the temporal location t.

4 Functional models

4.1 Deterministic models
1. A simple example
2. Endogenous and exogenous variables
3. Functions and rules
4. Functional models as process frames
5. Formal definition of functional models
6. Possible values and equivalent models
7. Elementary and derived functions
8. Set-valued functions
9. Dependencies between variables
10. Predictions and effective conditions
11. Multidimensional variables and constraints
12. Conditions for independence

4.2 Models with stochastic variables
1. Deterministic and stochastic functions
2. A simple example
3. Dependent and independent variables
4. Notations for stochastic models
5. Closed stochastic models
6. Concatenation of functions
7. Stochastically equivalent models
8. Direct and indirect predictions
9. Inversion of stochastic functions

4.3 Exogenous and unobserved variables
1. Stochastic exogenous variables?
2. Pseudo-indeterministic models
3. Defined unobserved variables
4. Endogenous unobserved variables
5. Assumptions about distributions
6. Interpretation of residuals

Statistical social research is based on data, values of statistical and relational variables, resulting from observations or interviews. However, rarely are data of interest on their own. The interest most often concerns more general questions which, in some sense, transcend the given data. Two kinds of generalizations must then be distinguished.

One idea, often discussed in the statistical literature, takes the cases for which data are available to be a subset of a larger population of cases and intends a generalization to that population. The conceptual framework of statistical variables allows an explicit formulation. Suppose a statistical variable, say $X : \Omega \longrightarrow \mathcal{X}$, represents the available data. This allows statistical statements about the reference set Ω, basically statements about $\mathrm{P}[X]$, the distribution of X in Ω. The reference set of the data, Ω, is then considered as a subset of a larger population $\Omega^\dagger$ for which an analogously defined variable, say $X^\dagger : \Omega^\dagger \longrightarrow \mathcal{X}$, can be assumed. The goal of the generalization is a statement about $\mathrm{P}[X^\dagger]$, the distribution of $X^\dagger$ in the population $\Omega^\dagger$, or some quantity derived from that distribution. For example, under certain circumstances it might be reasonable to believe that $\mathrm{P}[X^\dagger] \approx \mathrm{P}[X]$. In any case, the result is a descriptive statement about the distribution of $X^\dagger$ in the population for which the generalization is desired; the approach will therefore be called *descriptive generalization*.

While often reasonable, a descriptive approach to generalization has, in fact, severe limitations. The most important limitation results from the notion of a population. In order to be used as a reference set for a statistical variable, the elements of a population can only represent cases which actually do exist or have existed in the past.[1] However, interest in future possibilities often provides the main reason for an interest in generalizations. For example, one might be interested in the behavior of car drivers approaching traffic lights. Will they stop when the traffic light shows ‘red’? This question no longer refers to a definite set of realized facts but is a modal question that concerns the dependency of possibilities on conditions. Such questions cannot be answered in the conceptual framework of descriptive generalizations but require a different kind of generalization that will be called *modal generalization*.

The main linguistic tool for the formulation of modal generalizations are rules. Corresponding to different kinds of modal generalizations there are different kinds of rules. A rule might say, for example, what, in a situation of a certain kind, might happen, or will probably happen, or should be done, or can be achieved by performing some specified action. Scientific research is often concerned with rules that allow predictions, in particular with causal rules that are intended to show how facts, or events, of some kind depend on other facts and/or events.[2] Modal generalizations may then be called *predictive* or *causal generalizations*.

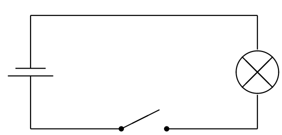
A widespread approach to predictive and causal generalizations consists in the construction of *functional models*, that is, models which show how one or more endogenous variables depend on one or more exogenous variables. The important point is that these variables, contrary to statistical variables, do not represent realized facts, but are intended to serve modal thinking about dependencies between possible facts and/or events. These variables will therefore be called *modal variables* and a specific notation will be used. They will be marked by a single point if stochastic or by double points if deterministic.

The present chapter presents an abstract discussion of functional models which are based on modal variables. It will be supposed that the interest is in predictive and causal generalizations. Models are therefore considered, not as descriptive, but as analytical models. The first section deals with deterministic functional models. The second section considers functional models which also contain stochastic variables and stochastic functions. The third section discusses some conceptual problems related to speculations about unobserved variables. Additional questions concerning the interpretation and application of functional models will be discussed in subsequent chapters.

4.1 Deterministic models

1. A simple example

A *functional model* consists of modal variables which are connected by functions. This section deals with *deterministic* functional models containing only deterministic variables connected by deterministic functions. The following diagram illustrates a simple example:



A battery and a bulb are connected by a circuit that can be closed or opened with a switch. Depending on the position of the switch the bulb gives light or not. A functional model for this situation uses three variables:[3] $\ddot{Y}$ records whether the bulb gives light ($\ddot{Y} = 1$) or not ($\ddot{Y} = 0$), $\ddot{X}$ records whether the switch is closed ($\ddot{X} = 1$) or not ($\ddot{X} = 0$), and $\ddot{Z}$ records whether the battery can provide power ($\ddot{Z} = 1$) or not ($\ddot{Z} = 0$). In addition there is a function that shows how the possible values of $\ddot{Y}$ depend on values of $\ddot{X}$ and $\ddot{Z}$. The following table defines the function:

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

(4.1)

The bulb gives light if the battery provides power and the switch is closed; in all other cases the bulb is off.

2. Endogenous and exogenous variables

A main ingredient of functional models are the *modal variables*.[4] Their definition requires:

(a) The *name of the variable* and an explication of its intended meaning. For example, a variable is called $\ddot{X}$ and it is intended to record whether a switch is open or closed.
(b) The *range* of the variable, that is, the set of its possible values must be defined and the intended meaning of the values must be explained.[5] For example, $\ddot{X}$ has the range $\{0, 1\}$ and 0 means that the switch is open and 1 means that it is closed.
(c) How modal variables get specific values.

The requirement (c) is particularly important in order to make functional models intelligible. A basic distinction between endogenous and exogenous variables must be made. *Endogenous modal variables* get their values as functions of other variables. In the example in § 1, $\ddot{Y}$ is an endogenous variable that gets its values as a function of the values of $\ddot{X}$ and $\ddot{Z}$. Modal variables which are not endogenous are called *exogenous modal variables*; they get their values from arbitrary assumptions. In the example $\ddot{X}$ and $\ddot{Z}$ are exogenous variables.

3. Functions and rules

The variables of a functional model are connected by functions. In a deterministic model the functions are of the form $f : \tilde{\mathcal{X}} \longrightarrow \tilde{\mathcal{Y}}$. The set $\tilde{\mathcal{X}}$, being the range of a variable $\ddot{X}$, is the function's domain, and $\tilde{\mathcal{Y}}$, being the range of another variable $\ddot{Y}$, is the function's codomain. The function assigns to each element $x \in \tilde{\mathcal{X}}$ exactly one element $f(x) \in \tilde{\mathcal{Y}}$. In this sense, given a value of $\ddot{X}$, the function uniquely determines a value of $\ddot{Y}$.[6]

Such dependency relations can be graphically depicted by arrows. In the just used example: $\ddot{X} \longrightarrow \ddot{Y}$. A variable can, of course, depend on two or more other variables simultaneously. For the example in § 1 one can use the following diagram: $\ddot{X} \longrightarrow \ddot{Y} \longleftarrow \ddot{Z}$. Note that the two arrows do not correspond to two separate functions; rather all arrows which lead to the same variable belong to one function. In this example, this is the function $g : \tilde{\mathcal{X}} \times \tilde{\mathcal{Z}} \longrightarrow \tilde{\mathcal{Y}}$ which assigns to each element $(x, z) \in \tilde{\mathcal{X}} \times \tilde{\mathcal{Z}}$ a value $g(x, z) \in \tilde{\mathcal{Y}}$ (as shown in (4.1)). Therefore, if two variables are connected by an arrow, this only shows that the model contains a function having one of the variables as its depending variable and the other one as an argument variable.

Since mathematical functions can be used for many different purposes one can reasonably ask for their specific meaning when used for the specification of functional models. Our discussion is based on the idea that they formulate *modeling rules* which show how values of the endogenous variables of a model are determined from values of other variables of the model. One might say that the functions express, or characterize, the processes that produce values of the endogenous variables of a model. Of course, such modeling rules are only valid for the model and they do not have any implications for processes in the real world. For example, referring to the model in § 1, it is a valid rule that the bulb gives light if the battery provides power and the switch is closed. But whether this will also be true of a real system that one has created by connecting a battery, a bulb, and a switch, is obviously a quite different question.

4. Functional models as process frames

A functional model can also be viewed as a process frame (see Section 3.1, § 4) that specifies how, beginning with values of the exogenous variables, values of the endogenous variables successively are produced. In the graphical presentation of the model the arrows depict the processes. If the direction of the arrows (functions) corresponds to a temporal ordering one can speak of a dynamic functional model. In any case a functional model permits the thinking of repeatable processes.

5. Formal definition of functional models

A functional model $\mathcal{M}$ can formally be defined as $\mathcal{M} = (\mathcal{V}, \mathcal{F})$ with $\mathcal{V}$ being a set of modal variables, and $\mathcal{F}$ being a set of functions which connect the variables in $\mathcal{V}$. A variable can occur in at most one function as a dependent variable; if it does it is called an endogenous variable, otherwise an exogenous variable. It is required that each member of $\mathcal{V}$ occurs in at least one function, either as an argument or as a dependent variable. A further requirement is that all argument variables are effective conditions for the corresponding dependent variables.[7]

To each functional model $\mathcal{M} = (\mathcal{V}, \mathcal{F})$ can be associated a directed graph $\mathcal{G}(\mathcal{M}) = (\mathcal{V}, \mathcal{K})$ with $\mathcal{V}$ the set of nodes and $\mathcal{K}$ the set of edges. The nodes correspond to the variables of the model, and if $V, V' \in \mathcal{V}$, there is a directed edge (an arrow) from V to V' if $\mathcal{F}$ contains a function having V as an argument variable and V' as the dependent variable.

The graph of a functional model shows the structure of the model. These graphs are always simple graphs without loops. If not explicitly said otherwise, it will be assumed that the graphs are weakly connected and do not contain cycles. Then, for example, the following model structures are possible with three variables (ignoring permutations of variable names)

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

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

Since it is possible that different models have the same structure, the following definition is often helpful: Two models, $\mathcal{M} = (\mathcal{V}, \mathcal{F})$ and $\mathcal{M}' = (\mathcal{V}', \mathcal{F}')$, are *structurally identical* if their graphs are isomorphic.

6. Possible values and equivalent models

The range of a variable specifies its possible values. However, which values are actually possible also depends on whether two ore more variables are considered simultaneously. The following notation is helpful: If V is a list of variables defined in a model $\mathcal{M} = (\mathcal{V}, \mathcal{F})$ (or a set of variables for which some ordering is given), $\mathrm{B}_{\mathcal{M}}[V]$ denotes the set of possible values of the variables in V within the model $\mathcal{M}$. For example, referring to the model in § 1,

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

but

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

is only a subset of $\tilde{\mathcal{X}} \times \tilde{\mathcal{Y}} \times \tilde{\mathcal{Z}}$.

The notion of possible values of sets of variables can be used to define equivalent models: Two models, $\mathcal{M} = (\mathcal{V}, \mathcal{F})$ and $\mathcal{M}' = (\mathcal{V}', \mathcal{F}')$, are *equivalent* if $\mathrm{B}_{\mathcal{M}}[\mathcal{V}] = \mathrm{B}_{\mathcal{M}'}[\mathcal{V}']$. Equivalence obviously implies a one-to-one mapping from $\mathcal{V}$ to $\mathcal{V}'$; it is not implied, however, that the models are structurally identical. This is shown, for example, by the following models:

$$\ddot{X} \longrightarrow \ddot{Y} \longrightarrow \ddot{Z} \quad \text{and} \quad \ddot{X} \longleftarrow \ddot{Y} \longrightarrow \ddot{Z}$$

Both models will be equivalent if one specifies $\tilde{\mathcal{X}} := \{0,1\}$, $\tilde{\mathcal{Y}} := \{0,2\}$, $\tilde{\mathcal{Z}} := \{0,8\}$, and assumes the functions $y = 2x$ and $z = 4y$ for the first model and the functions $x = 0.5y$ and $z = 4y$ for the second model.

7. Elementary and derived functions

The definition of a functional model provides for each endogenous variable a function that shows how values of the variable depend on values of other variables. These functions (the elements of $\mathcal{F}$) will be called *elementary functions*. Other functions can often be derived from the elementary functions. For example, if a model has the structure

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

one can derive a function $f_{yz} : \tilde{\mathcal{Z}} \longrightarrow \tilde{\mathcal{Y}}$ that concatenates the two elementary functions: $f_{yz}(z) = f_{yx}(f_{xz}(z))$. As another example with a somewhat more complicated structure consider the model

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

with elementary functions $f_{yxz}(x,z)$ and $f_{zx}(x)$. From these one can derive a function $f_{yx}(x) = f_{yxz}(x, f_{zx}(x))$, corresponding to a single arrow from $\ddot{X}$ to $\ddot{Y}$, which no longer explicitly refers to values of $\ddot{Z}$. Moreover, the function only has an exogenous argument variable.

For each endogenous variable of a model, one can always derive a function which has only exogenous variables as arguments. This leads to the notion of a *reduced model*. Beginning with a functional model $\mathcal{M} = (\mathcal{V}, \mathcal{F})$, the reduced form of $\mathcal{M}$ only contains the exogenous and the ultimate endogenous variables (variables without a successor) together with the functions derived from the original model.

8. Set-valued functions

The elementary functions of a model are single-valued functions: these functions assign to each value of their domain exactly one value of their codomain. It is sometimes helpful to use instead *set-valued functions* which assign to each value of their domain a subset of their codomain. The example from § 1 can serve to illustrate the notion. Although there is no single-valued function $\ddot{Y} \longrightarrow (\ddot{X}, \ddot{Z})$, a set-valued function can be used to characterize the relationship:

$$y = 0 \longrightarrow \{(0,0), (0,1), (1,0)\}$$
$$y = 1 \longrightarrow \{(1,1)\}$$

An extension of the notation introduced in § 6 allows flexible references to set-valued functions which can be derived from the elementary functions of a model: $\mathrm{B}_{\mathcal{M}}[\ddot{Y}|\ddot{X}=x]$ denotes the set of all values of $\ddot{Y}$ which are possible in the model $\mathcal{M}$ given the condition $\ddot{X}=x$. Similarly, several variables can be used before as well as behind the condition symbol. For example, if $\mathcal{M}$ refers to the model in § 1, then $\mathrm{B}_{\mathcal{M}}[\ddot{X},\ddot{Z}|\ddot{Y}=0] = \{(0,0),(0,1),(1,0)\}$, $\mathrm{B}_{\mathcal{M}}[\ddot{X},\ddot{Z}|\ddot{Y}=1] = \{(1,1)\}$, and the set-valued function $\ddot{Y} \longrightarrow (\ddot{X}, \ddot{Z})$ can be written as $y \longrightarrow \mathrm{B}_{\mathcal{M}}[\ddot{X},\ddot{Z} \,|\, \ddot{Y}=y]$.

9. Dependencies between variables

When investigating relationships between variables it is useful to distinguish two kinds of dependencies. The following definitions refer to a model $\mathcal{M} = (\mathcal{V}, \mathcal{F})$ and two variables $\ddot{X}$ and $\ddot{Y}$ in $\mathcal{V}$: $\ddot{Y}$ *depends* on $\ddot{X}$ if $\mathrm{B}_{\mathcal{M}}[\ddot{Y}|\ddot{X}=x] \neq \mathrm{B}_{\mathcal{M}}[\ddot{Y}|\ddot{X}=x']$ for some values $x, x' \in \mathrm{B}_{\mathcal{M}}[\ddot{X}]$. $\ddot{Y}$ *functionally depends* on $\ddot{X}$ if $\ddot{Y}$ depends on $\ddot{X}$ and there is a directed path leading from $\ddot{X}$ to $\ddot{Y}$. In this case it is also said that $\ddot{X}$ is an *effective condition* for $\ddot{Y}$.

Dependence (in contrast to functional dependence) is in most applications symmetrical: if $\ddot{Y}$ depends on $\ddot{X}$ then also $\ddot{X}$ depends on $\ddot{Y}$.[8] This allows the definition of a symmetric notion of independence: With respect to some model $\mathcal{M}$, two variables $\ddot{X}$ and $\ddot{Y}$ are *independent*, symbolically denoted by $\ddot{X} \perp \ddot{Y}$, if neither $\ddot{X}$ depends on $\ddot{Y}$ nor $\ddot{Y}$ depends on $\ddot{X}$. Accordingly, the variables are called *dependent* (in a symmetrical sense) if they are not independent. For example, with respect to the model in § 1, the variables $\ddot{X}$ and $\ddot{Z}$ are independent, but $\ddot{X}$ and $\ddot{Y}$ as well as $\ddot{Z}$ and $\ddot{Y}$ are dependent.

If two variables are independent it obviously follows that none can be functionally dependent on the other one. On the other hand, if two variables are not functionally dependent, it does not follow that they are independent; it is quite possible that both variables depend on a third one. Note also that from pairwise independence of three or more variables it does not follow that the variables are simultaneously independent. For example, if $\ddot{X} \perp \ddot{Y}$ and $\ddot{X}' \perp \ddot{Y}$, one cannot conclude that $(\ddot{X}, \ddot{X}') \perp \ddot{Y}$.[9]

A further useful notion is *conditional independence*. If $\ddot{X}$, $\ddot{Y}$, and $\ddot{Z}$ are variables of some model $\mathcal{M}$, two versions of this notion can be defined as follows: $\ddot{X}$ and $\ddot{Y}$ are *independent conditional on* $\ddot{Z} = z$, symbolically denoted by $\ddot{X} \perp \ddot{Y} | \ddot{Z} = z$, if $\mathrm{B}_{\mathcal{M}}[\ddot{Y} | \ddot{X} = x, \ddot{Z} = z] = \mathrm{B}_{\mathcal{M}}[\ddot{Y} | \ddot{X} = x', \ddot{Z} = z]$ for all values $x, x' \in \mathrm{B}_{\mathcal{M}}[\ddot{X} | \ddot{Z} = z]$. $\ddot{X}$ and $\ddot{Y}$ are *independent conditional on* $\ddot{Z}$, symbolically denoted by $\ddot{X} \perp \ddot{Y} | \ddot{Z}$, if $\ddot{X} \perp \ddot{Y} | \ddot{Z} = z$ for all $z \in \mathrm{B}_{\mathcal{M}}[\ddot{Z}]$. Referring to the model in § 1 one finds, for example, $\ddot{X} \perp \ddot{Y} | \ddot{Z} = 0$, but not $\ddot{X} \perp \ddot{Y} | \ddot{Z} = 1$.

Remarkably, conditional independence does not imply independence. As an illustration consider a model $\ddot{X} \longleftarrow \ddot{Z} \longrightarrow \ddot{Y}$ where $\ddot{X}$ and $\ddot{Y}$ are functionally dependent on a variable $\ddot{Z}$. This implies $\ddot{X} \perp \ddot{Y} | \ddot{Z}$, but not $\ddot{X} \perp \ddot{Y}$.[10] Conversely, as shown by the example in footnote 9, $\ddot{X} \perp \ddot{Y}$ does not imply $\ddot{X} \perp \ddot{Y} \,|\, \ddot{Z}$.

10. Predictions and effective conditions

If two variables of a model $\mathcal{M}$, say $\ddot{X}$ and $\ddot{Y}$, are dependent, each variable can be used to predict values of the other one. Two forms can be distinguished: Given $\ddot{X} = x$, $\ddot{Y}$ can be *uniquely predicted* if the set $\mathrm{B}_{\mathcal{M}}[\ddot{Y} | \ddot{X} = x]$ contains a single element. Given $\ddot{X} = x$, $\ddot{Y}$ can be *indeterminately predicted* if the set $\mathrm{B}_{\mathcal{M}}[\ddot{Y} | \ddot{X} = x]$ contains two or more elements but is a proper subset of $\mathrm{B}_{\mathcal{M}}[\ddot{Y}]$.

Using these definitions, predicting values of a variable $\ddot{Y}$ from the knowledge of the value of a variable $\ddot{X}$ does not require that $\ddot{Y}$ functionally depends on $\ddot{X}$. For example, referring to the model in § 1, given $\ddot{Y} = 1$ one can uniquely predict values of $\ddot{X}$ and $\ddot{Z}$; but, of course, $\ddot{Y}$ is an effective condition neither for $\ddot{X}$ nor for $\ddot{Z}$.

11. Multidimensional variables and constraints

Variables of a functional model can be multi-dimensional. For example, one can assume that a variable $\ddot{X}$ consists of m components, having the form $\ddot{X} = (\ddot{X}_1, \ldots, \ddot{X}_m)$. In the same way one can combine already defined variables into a single multi-dimensional variable; for example, given variables $\ddot{X}$ and $\ddot{Y}$ with ranges $\tilde{\mathcal{X}}$ and $\tilde{\mathcal{Y}}$, respectively, they can be combined into a two-dimensional variable $(\ddot{X}, \ddot{Y})$ with the range $\tilde{\mathcal{X}} \times \tilde{\mathcal{Y}}$.

Multi-dimensional variables are particularly useful in the definition of constraints on their ranges. For example, if $\ddot{X}$ and $\ddot{Y}$ are two variables with ranges $\tilde{\mathcal{X}} = \tilde{\mathcal{Y}} = \{1, \ldots, 5\}$, it might be required to restrict the possible joint range by the constraint $x + y \leq 5$. The variables are then dependent, but their dependency cannot be expressed by functions.[11] However, the condition that creates the dependency can easily be captured by referring to the two-dimensional variable $(\ddot{X}, \ddot{Y})$ and specifying the range by $\{(x,y) \in \tilde{\mathcal{X}} \times \tilde{\mathcal{Y}} \,|\, x+y \leq 5\}$.

It is evident that constraints can also be defined for two or more exogenous variables of a model. This has the important implication that exogenous variables are independent only if no constraints exist. Of course, they are never functionally dependent because only endogenous variables of a model can functionally depend on other variables.

12. Conditions for independence

The possible existence of constraints is of particular importance for the question whether dependent variables can be made independent by proper conditions. This is always possible in deterministic models without constraints: If $\ddot{X}$ and $\ddot{Y}$ are two dependent variables (by assumption also endogenous and not functionally dependent), and the set of preceding variables (variables from which a directed path leads to $\ddot{X}$ or $\ddot{Y}$) consists of $\ddot{Z}_1, \ldots, \ddot{Z}_m$, the conditional independence $\ddot{X} \perp \ddot{Y} \,|\, \ddot{Z}_1, \ldots, \ddot{Z}_m$ will always be true.

This statement, sometimes called the *deterministic Markov condition*, is not generally true if there are constraints. It is quite possible, then, that no variables exist that can be used as conditions in order to make $\ddot{X}$ and $\ddot{Y}$ conditionally independent. Of course, one can always construct a variable $\ddot{Z}$ such that $\ddot{X} \perp \ddot{Y} \,|\, \ddot{Z}$.[12] But this is then a fictitious variable not contained in the actual model.

The argument implies a qualification for what has been called *the principle of common dependence*: If two variables of a functional model $\mathcal{M}$, say $\ddot{X}$ and $\ddot{Y}$, are dependent, then either $\ddot{X}$ is functionally dependent on $\ddot{Y}$, or $\ddot{Y}$ is functionally dependent on $\ddot{X}$, or there exists a further variable $\ddot{Z}$ such that both $\ddot{X}$ and $\ddot{Y}$ are functionally dependent on $\ddot{Z}$ and $\ddot{X} \perp \ddot{Y} \,|\, \ddot{Z}$. While it is always possible to construct such a variable $\ddot{Z}$, if a model contains constraints it is not guaranteed that such a variable exists as part of the model.

4.2 Models with stochastic variables

This section deals with functional models which contain not only deterministic variables, denoted $\ddot{X}$, $\ddot{Y}$, and so on, but also stochastic variables that will be distinguished from deterministic variables by a single dot, e.g. $\dot{X}$, $\dot{Y}$, and so on.

1. Deterministic and stochastic functions

If a functional model contains not only deterministic but also stochastic variables two kinds of functions must be distinguished:

(a) *Deterministic functions* which can connect deterministic variables $\ddot{X}$ and $\ddot{Y}$, or stochastic variables $\dot{X}$ and $\dot{Y}$, in the following way: $\ddot{X} \longrightarrow \ddot{Y}$ or $\dot{X} \longrightarrow \dot{Y}$. In both cases the deterministic function assigns to each element in $\tilde{\mathcal{X}}$, the range of $\ddot{X}$ or $\dot{X}$, exactly one element in $\tilde{\mathcal{Y}}$, the range of $\ddot{Y}$ or $\dot{Y}$.[13]

In the second case one can derive the distribution of $\dot{Y}$ from the distribution of $\dot{X}$.[14]

(b) *Stochastic functions* which can connect variables in the following way: $\ddot{X} \longrightarrow\!\!\!\!\rightarrow \dot{Y}$ or $\dot{X} \longrightarrow\!\!\!\!\rightarrow \dot{Y}$. In both cases the stochastic function assigns to each element in $\tilde{\mathcal{X}}$, the range of $\ddot{X}$ or $\dot{X}$, a conditional distribution, $\Pr[\dot{Y}|\ddot{X}=x]$ or $\Pr[\dot{Y}|\dot{X}=x]$, respectively.[15]

Because these are different kinds of functions they are distinguished also graphically: $\longrightarrow$ is used for deterministic functions and $\longrightarrow\!\!\!\!\rightarrow$ is used for stochastic functions.

2. A simple example

A simple model for a toaster can serve to illustrate the notion of a stochastic function. There are two variables. The deterministic variable $\ddot{X}$, with range $\tilde{\mathcal{X}}$, records the position of a lever that can be used to adjust the toasting duration, and the stochastic variable $\dot{Y}$, with range $\tilde{\mathcal{Y}}$, records the realized toasting duration. A stochastic function $\ddot{X} \longrightarrow\!\!\!\!\rightarrow \dot{Y}$ connects the two variables and provides, for each value $x \in \tilde{\mathcal{X}}$, a conditional distribution $\Pr[\dot{Y}|\ddot{X}=x]$. The function shows how the distribution of toasting durations depends on the position of the lever. Of course, it might suffice to consider a function $x \longrightarrow \mathrm{E}(\dot{Y}|\ddot{X}=x)$ if one is interested only in mean toasting durations.

3. Dependent and independent variables

If stochastic variables are involved dependency relations can be defined in two different ways. The following definitions refer to a model $\mathcal{M} = (\mathcal{V}, \mathcal{F})$ and two variables $\ddot{X}$ and $\dot{Y}$ (analogous definitions apply if a stochastic variable $\dot{X}$ is used instead of the deterministic variable $\ddot{X}$).

- One can apply the definitions of § 9 in Section 4.1, in particular: $\ddot{X}$ and $\dot{Y}$ are *independent*, symbolically denoted by $\ddot{X} \perp \dot{Y}$, if $\mathrm{B}_{\mathcal{M}}[\dot{Y}|\ddot{X}=x] = \mathrm{B}_{\mathcal{M}}[\dot{Y}|\ddot{X}=x']$ for all values $x, x' \in \mathrm{B}_{\mathcal{M}}[\ddot{X}]$; otherwise the variables are *dependent*.
- One can exploit the fact that at least one of the variables is stochastic and use the following definition: $\ddot{X}$ and $\dot{Y}$ are *stochastically independent*, symbolically denoted by $\ddot{X} \perp\!\!\!\perp \dot{Y}$, if $\Pr[\dot{Y}|\ddot{X}=x] = \Pr[\dot{Y}|\ddot{X}=x']$ for all values $x, x' \in \mathrm{B}_{\mathcal{M}}[\ddot{X}]$; otherwise the variables are *stochastically dependent*.[16] Stochastic independence can be defined equivalently in the following way:

$$\text{For all } (x,y) \in \mathrm{B}_{\mathcal{M}}[\ddot{X}, \dot{Y}]\text{:}\;\; \Pr(\dot{Y}=y\,|\,\ddot{X}=x) = \Pr(\dot{Y}=y)$$

and, if both variables are stochastic:

$$\text{For all } (x,y) \in \mathrm{B}_{\mathcal{M}}[\dot{X}, \dot{Y}]\text{:}\;\; \Pr(\dot{Y}=y, \dot{X}=x) = \Pr(\dot{Y}=y)\Pr(\dot{X}=x)$$

Both notions of dependence and independence are symmetric,[17] and the following statements can be derived: If two variables are dependent they are also stochastically dependent; and conversely: if they are stochastically independent they are also independent. It is quite possible, however, that two independent variables are stochastically dependent. As in the deterministic case, a conditional independence $\dot{X} \perp\!\!\!\perp \dot{Y} | \dot{Z}$ does not imply the simple independence $\dot{X} \perp\!\!\!\perp \dot{Y}$, and vice versa; and pairwise stochastic independence of three or more variables does not imply simultaneous stochastic independence of the variables.

4. Notations for stochastic models

The notation $\mathcal{M} = (\mathcal{V}, \mathcal{F})$, introduced in Section 4.1 for deterministic models, can also be used for stochastic models. The specification of $\mathcal{V}$ must then distinguish between deterministic and stochastic variables, and the specification of $\mathcal{F}$ must distinguish between deterministic and stochastic functions. It will be assumed that a stochastic model contains at least one endogenous stochastic variable. This implies that there is also at least one stochastic function.

Exogenous variables can be deterministic or stochastic. If a model contains exogenous stochastic variables assumptions about their distributions are taken as being a part of the model specification. Of course, if there are two or more exogenous stochastic variables their common distribution must be specified. It will furthermore be assumed that there are no constraints connecting stochastic and deterministic exogenous variables. This implies that stochastic exogenous variables are always stochastically independent of deterministic exogenous variables.[18]

Like deterministic models, stochastic models can be represented by directed graphs. Each node refers to a deterministic or stochastic variable, and each edge refers to a deterministic ($\longrightarrow$) or stochastic ($\longrightarrow\!\!\!\!\rightarrow$) function (of course, edges belonging to the same function are of the same form). So one can again use the following definition: Two models $\mathcal{M} = (\mathcal{V}, \mathcal{F})$ and $\mathcal{M}' = (\mathcal{V}', \mathcal{F}')$ are *structurally identical* if their graphs are isomorphic.

5. Closed stochastic models

The general definition of § 4 permits stochastic models in which all exogenous (and consequently all endogenous) variables are stochastic. Such models will be called *closed stochastic models*. A simple example is

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

In closed stochastic models, the joint distribution of all variables can be derived easily:

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

In particular, the joint distribution of all endogenous variables of the model is just the marginal distribution of these variables; in the example:

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

Authors who discuss stochastic models very often use closed models right from the beginning and (consequently) presuppose the existence of a common distribution of all variables defined in the model.[19] This is possibly due to a widespread belief that stochastic models are to be understood as (sampling) models of statistical data. However, there is no direct relationship between functional models and statistical data (this will be further discussed in Chapter 6). In fact, as will be argued in Section 4.3, there is no reason why stochastic exogenous variables should be required in the definition of analytical functional models.

6. Concatenation of functions

Functions along a directed path between variables can be concatenated. With three variables, the following elementary forms are possible:

(1) $\dot{Z} \xrightarrow{g} \dot{X} \xrightarrow{f} \dot{Y}$. The concatenation is $y = f(g(z))$.
(2) $\dot{Z} \xrightarrow{g} \dot{X} \twoheadrightarrow \dot{Y}$. The concatenation is

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

(3) $\dot{Z} \twoheadrightarrow \dot{X} \xrightarrow{f} \dot{Y}$. The concatenation is

$$z \longrightarrow \Pr(\dot{Y} = y \,|\, \dot{Z} = z) = \textstyle\sum_{x \in f^{-1}(\{y\})} \Pr(\dot{X} = x \,|\, \dot{Z} = z)$$

(4) $\dot{Z} \twoheadrightarrow \dot{X} \twoheadrightarrow \dot{Y}$. A stochastic function that directly connects $\dot{Z}$ and $\dot{Y}$ has the form:

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

One might ask whether $\dot{Y}$ and $\dot{Z}$ are stochastically independent conditional on $\dot{X}$. While this is not, in general, true if the three variables have an arbitrary common distribution, it can safely be assumed for the model structures shown above. In the first three cases the independence can directly be proved. In the last case it follows from the omission of an arrow leading from $\dot{Z}$ to $\dot{Y}$. This also allows the simplification of the stochastic function:

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

7. Stochastically equivalent models

Although the definition of equivalent models introduced in Section 4.1 (§ 6) is applicable to stochastic models as well it is rarely useful. A stochastic version of equivalence is often more appropriate. This version is based on the notion of a *reduced model* already defined for deterministic models in § 7 of Section 4.1. The definition for a stochastic model $\mathcal{M} = (\mathcal{V}, \mathcal{F})$ proceeds in two steps: First one eliminates all intermediate variables leading to a model that only contains exogenous and ultimate endogenous variables. Then, if present, one also eliminates stochastic exogenous variables. This can be done by mixing the previously derived functions with the distributions of the stochastic exogenous variables (which are known as part of the model specification). The final result is a reduced model of the form $\ddot{X} \longrightarrow\!\!\!\!\rightarrow \dot{Y}$ where $\ddot{X}$ refers to all deterministic exogenous variables and $\dot{Y}$ refers to all ultimate endogenous variables.[20]

This notion of a reduced model can now be used for the following definition of stochastic equivalence: Two stochastic models are *stochastically equivalent* if their reduced models are identical. The definition implies that each model is, in particular, stochastically equivalent with its reduced model. A model of the form $\ddot{X} \longrightarrow \dot{Y} \longleftarrow \dot{Z}$ can serve as an example. The reduced form is $\ddot{X} \longrightarrow\!\!\!\!\rightarrow \dot{Y}$. Although the models are formally and structurally different they are stochastically equivalent.

8. Direct and indirect predictions

Let $\mathcal{M} = (\mathcal{V}, \mathcal{F})$ denote a deterministic model. All elementary functions (the elements of $\mathcal{F}$) and also all functions which can be recursively derived by concatenations will be called *direct prediction functions*. For example, the model

$$\ddot{X} \longrightarrow \ddot{Y} \longleftarrow \ddot{Z} \longleftarrow \ddot{V} \tag{4.2}$$

allows consideration of three direct prediction functions: two elementary functions, and a derived function for $\ddot{Y}$ with arguments $\ddot{X}$ and $\ddot{V}$. In contrast, functions which can be used for unique or indeterminate predictions will be called *indirect prediction functions* if they cannot be derived by concatenating elementary functions.

As was discussed in Section 4.1, these functions must often be formulated as set-valued functions. For example, referring to (4.2), one could define a set-valued function $v \longrightarrow \mathrm{B}_{\mathcal{M}}[\ddot{X} \mid \ddot{V} = v]$ that connects the variables $\ddot{V}$ and $\ddot{X}$, and this would be an indirect prediction function if $\mathrm{B}_{\mathcal{M}}[\ddot{X} \mid \ddot{V} = v]$ is a proper subset of $\mathrm{B}_{\mathcal{M}}[\ddot{X}]$ for at least one value v.

An analogous distinction between direct and indirect prediction functions is possible for stochastic models. As an example that parallels (4.2) consider the model

$$\ddot{X} \longrightarrow\!\!\!\!\rightarrow \dot{Y} \twoheadleftarrow\!\!\!\!-\!\!- \dot{Z} \twoheadleftarrow\!\!\!\!-\!\!- \ddot{V}$$

There are again three direct, now stochastic prediction functions; and it would also be possible to construct set-valued functions. However, set-valued functions derived from stochastic models are rarely useful for predictions. We therefore briefly consider the question how stochastic versions of indirect prediction functions might be constructed.

9. Inversion of stochastic functions

Consider as an example the stochastic function $\dot{X} \longrightarrow\!\!\!\!\rightarrow \dot{Y}$. Without further preconditions it will not be possible to derive a stochastic function $\dot{Y} \longrightarrow\!\!\!\!\rightarrow \dot{X}$ which would allow the use of values of $\dot{Y}$ to predict conditional distributions of $\dot{X}$. The specification of $\Pr[\dot{X}]$, the marginal distribution of $\dot{X}$, is needed. Given the distribution of $\dot{X}$, the joint distribution of $\dot{X}$ and $\dot{Y}$ is

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

from which the stochastic function $\dot{Y} \longrightarrow\!\!\!\!\rightarrow \dot{X}$ could be derived.

Except for the special (and problematic) case of closed models, marginal distributions cannot be derived from the original model. The crucial question therefore is how one can find, and justify, enlarged models. A simple example can serve to illustrate the problem. The model is $\ddot{X} \longrightarrow\!\!\!\!\rightarrow \dot{Y}$ with binary variables as follows: $\ddot{X} = 1$ if a patient was given a particular therapy, $\ddot{X} = 0$ otherwise, and $\dot{Y} = 1$ if there were a successful recovery and $\dot{Y} = 0$ otherwise. The stochastic function is given by

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

Now consider an indirect prediction problem: Knowing that the recovery was successful, how can one predict whether the therapy was applied or not? Obviously, neither a unique nor an indeterminate (set-valued) prediction is possible.

This may motivate a probabilistic approach. The formal requirement is to substitute the deterministic variable $\ddot{X}$ by a stochastic variable $\dot{X}$ (having the same range). This allows one to consider a different version of the indirect prediction problem: Given knowledge about a value of $\dot{Y}$, what is the conditional distribution of $\dot{X}$? For the example assume that one knows $\dot{Y}=1$. This is, however, insufficient to compute a conditional distribution of $\dot{X}$, one also needs an assumption about the unconditional distribution of $\dot{X}$. For example, the assumption could be that both possibilities have equal probability: $\Pr(\dot{X}=1) = 0.5$. Using this information about the distribution of $\dot{X}$ together with the value of $\dot{Y}$ it is possible to calculate the common distribution:

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

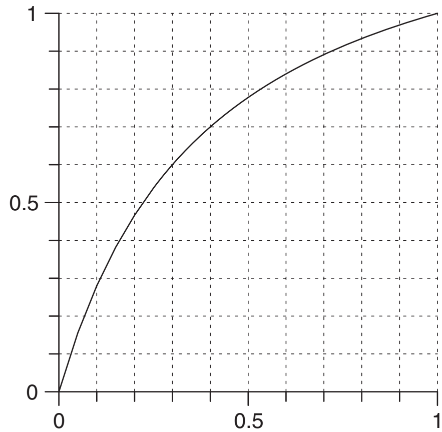


Figure 4.1 The graph shows how $\Pr(\dot{X}=1\,|\,\dot{Y}=1)$, given on the Y-axis, depends on assumptions about $\Pr(\dot{X}=1)$, given on the X-axis.

and one can derive the following answer to the prediction problem:

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

Of course, as illustrated in Figure 4.1, the answer depends heavily on the unconditional distribution of $\dot{X}$.

4.3 Exogenous and unobserved variables

1. Stochastic exogenous variables?

A main task of functional models is to show how endogenous variables (values of deterministic or distributions of stochastic endogenous variables) depend *on values* of exogenous variables. The exogenous variables can be deterministic or stochastic. However, if exogenous variables are stochastic their distributions are actually irrelevant. Consider the following example:

$$\text{(a)}\quad \begin{array}{l} \ddot{X} \searrow \\ \qquad\quad \twoheadrightarrow \dot{Y} \\ \ddot{Z} \nearrow \end{array} \qquad \text{(b)}\quad \begin{array}{l} \ddot{X} \searrow \\ \qquad\quad \twoheadrightarrow \dot{Y} \\ \dot{Z} \nearrow \end{array} \tag{4.3}$$

The only difference is that version (a) uses a deterministic variable $\ddot{Z}$ and version (b) uses a stochastic variable $\dot{Z}$ instead. The stochastic functions that connect the variables are in both versions identical:

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

The distribution of $\dot{Z}$ which must be assumed for version (b) is obviously irrelevant for this function.

Consequently, if the goal is to learn about dependencies on *values* of exogenous variables it is reasonable to use deterministic variables that allow assumptions about their values. On the other hand, using stochastic exogenous variables seems sensible just in those cases where assumptions about specific values cannot be justified. Instead, one assumes distributions and this allows the elimination of variables in the stochastic functions of the model. For example, in version (b), one can derive the function

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

in which the dependence on $\dot{Z}$ no longer occurs.

However, if stochastic exogenous variables can be eliminated, and eventually must be eliminated in order to use a model for conditional predictions, why introduce such variables at all? A widespread idea is that such variables can be used to represent "unobserved variables." It is difficult, however, to give a clear meaning to this idea. Below, it will be suggested that this is only possible for explicitly defined unobserved variables. To prepare the discussion, the next paragraph briefly considers pseudo-indeterministic models which are widespread in the literature.

2. Pseudo-indeterministic models

Consider a simple stochastic model

$$\ddot{X} \twoheadrightarrow \dot{Y} \tag{4.4}$$

Instead, one can also consider a model of the form

$$\ddot{X} \longrightarrow \dot{Y} \longleftarrow \dot{Z} \tag{4.5}$$

which is stochastically equivalent to (4.4).[21] Both versions of the model suggest different interpretations. Version (4.4) requires a stochastic interpretation: the model shows how *the distribution* of $\dot{Y}$ depends on values of the exogenous variable $\ddot{X}$. Version (4.5), on the other hand, seems to allow a deterministic interpretation. In fact, this model presupposes a deterministic function $y=f(x,z)$ by which $\ddot{X}$ and $\dot{Z}$ generate *specific values* of $\dot{Y}$. Models having the form (4.5) are therefore often called *pseudo-indeterministic models*.[22]

In the literature most discussions of functional models start from pseudo-indeterministic models. The most prominent examples are regression models of the form

$$\dot{Y} = g(\ddot{X}) + \dot{Z} \tag{4.6}$$

with a deterministic function g. Such models often serve as the starting point for more complex definitions, including all kinds of structural equations models.[23] The crucial questions are in all cases the same: What meaning can be given to exogenous stochastic variables and their distributions? Can they be interpreted as "unobserved variables"? How can assumptions about their distributions be justified?[24]

3. Defined unobserved variables

The expression 'unobserved variables' can be used in two different ways. The expression may refer to explicitly defined variables for which data are missing; if this is meant we speak of *defined* unobserved variables. The expression might as well refer to stochastic exogenous variables of a pseudo-indeterministic model interpreted as "unknown factors."

No problems arise if defined unobserved variables are used for the definition of a functional model. With respect to the definition of a model there is actually no difference between observed and unobserved variables. The fact that no data are available for some variables only complicates the estimation of model parameters and, of course, may also restrict the applicability of a model. As an illustration, consider the model

$$\ddot{X} \longrightarrow\!\!\!\!\!\rightarrow \dot{Y} \leftarrow\!\!\!\!\!\longleftarrow \ddot{Z} \tag{4.7}$$

$\dot{Y}$ refers to a child's educational success (0 or 1), and $\ddot{X}$ records the school type (1 or 2). In addition there is an unobserved variable $\ddot{Z}$ for the parents' educational level (0 or 1). The model states the hypothesis that the child's educational success depends on the school type and on the parents' educational level.

Whether variables are observed or unobserved obviously is of no importance for the definition of a model. In particular, there is no reason to treat unobserved exogenous variables as stochastic variables. This is only required if estimates of model parameters are used for predictions.

If data are only available for $\ddot{X}$ and $\dot{Y}$ it is not possible to estimate (4.7); rather, only a model of the form (4.4) can be estimated. This also provides a reason for substituting the deterministic variable $\ddot{Z}$ by a stochastic variable $\dot{Z}$: The assumption of a distribution for $\dot{Z}$ allows one to think of (4.4) as being a reduced form of the original model (4.7). Of course, a distributional assumption is no substitute for missing data; the assumption nevertheless permits a conceptual clarification of the relationship between the model that can be estimated with the available data and the model that formulates the theoretical hypothesis. In the example, the theoretical model suggests that the relationship between school type and educational success, as it is found in the available data, also depends on an unknown distribution of parents' educational levels. Note that this formulation requires the notion of a statistical distribution which presupposes a specified population. On the other hand, being part of a functional model, the variable $\dot{Z}$ does not refer to any specific population and its distribution is conceptually different from a statistical distribution.

4. Endogenous unobserved variables

Defined unobserved variables can also occur as endogenous variables of a model. To illustrate, the example of the previous paragraph is modified in the following way:

$$\begin{tikzcd} \ddot{Z} \arrow[d] \arrow[dr, two heads] & \\ \dot{X} \arrow[r, two heads] & \dot{Y} \end{tikzcd} \tag{4.8}$$

In addition to the hypothesis that the child's educational success depends on the school type and on the parents' educational level, there is now the further hypothesis that the selection of a school type depends on the parents' educational level.

Now assume that $\dot{X}$ is an unobserved variable and data are available only for the parent's educational level, $\ddot{Z}$, and the child's educational success, $\dot{Y}$. Again, the data only allows the estimation of a reduced model, $\ddot{Z} \longrightarrow\!\!\!\!\rightarrow \dot{Y}$, corresponding to the stochastic function

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

However, no assumption about the distribution of $\dot{X}$ is required. The reduced model can be derived from the original model without any additional assumptions.

5. Assumptions about distributions

The examples have shown that there is an important difference between exogenous and endogenous stochastic variables. Exogenous stochastic variables require assumptions about their distributions; in contrast, one neither directly nor indirectly makes assumptions about endogenous stochastic variables, instead, one specifies stochastic functions which show how distributions of these variables depend on values of other variables. Consequently, only conditional distributions are needed, but these are part of the model's definition.

Consider, for example, the variable $\dot{X}$ in model (4.8) which refers to the school type. Although it is a stochastic variable it is not possible to derive, or even to think of, just one distribution of this variable. The model only shows how the distribution of $\dot{X}$ depends on values of an exogenous variable $\ddot{Z}$, but makes no assumptions about these values. There also is no immediate relationship between the stochastic variable $\dot{X}$ and a statistical variable X which records the distribution of children's school types in some specified population. A functional model does not relate to any population, and its deterministic as well as stochastic variables are conceptually different from statistical variables.

Since exogenous stochastic variables require assumptions about their distributions, the question arises what meanings can be given to these distributions, and how it might be possible to justify assumptions. Possible answers depend on the reasons for using stochastic exogenous variables in the formulation of a functional model. As argued earlier, since these models are intended to show how endogenous variables depend on values of exogenous variables, stochastic exogenous variables are only reasonable if corresponding data are missing and it is therefore necessary to employ a reduced model.

Further considerations depend on whether the problem of missing data occurs when estimating a model or when using a model for predictions. In the first case one often needs assumptions about distributions of unobserved variables already for the formulation of model parameterizations that can be estimated with the available data. The second case is somewhat different. The model (4.7) can serve as an example. Assume that this model is to be used in order to predict, for some specific child (or some collection of children), how its educational success depends on the school type. If information about the parents' educational level were available the model could be applied immediately. If such information is not available one can either confine oneself to conditional statements, or one can presuppose a *subjective* probability distribution for the unobserved variable and then derive subjective expectations.

6. *Interpretation of residuals*

As was argued previously, no conceptual problems arise from defined unobserved variables. In contrast, stochastic exogenous variables, as used in the formulation of pseudo-indeterministic models, create conceptual problems already from the beginning. Consider, for example, the model (4.5). In order to justify the existence of a deterministic function $y = f(x, z)$ one would need to assume that the stochastic variable $\dot{Z}$ captures *all* conditions, except $\ddot{X}$, on which values of $\dot{Y}$ depend. It is therefore impossible to give this variable any empirically explicable meaning.

At first sight, it might seem possible to set up a model of the form

$$\ddot{X} \longrightarrow \dot{Y} \longleftarrow (\dot{Z}_1, \ldots, \dot{Z}_m) \tag{4.9}$$

with variables $\dot{Z}_1, \ldots, \dot{Z}_m$ intended to represent all conditions which are possibly relevant in addition to $\ddot{X}$. However, since no one can know all possibly relevant conditions, it is impossible to formulate such a model explicitly. But even if that would be possible the question still remains how to define a function

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

that maps the variables $\dot{Z}_1, \ldots, \dot{Z}_m$ into a single variable $\dot{Z}$, and how to justify a distributional assumption about $\dot{Z}$.

It is remarkable that such questions are never considered when pseudo-indeterministic models are used. Instead, $\dot{Z}$ is interpreted as a residual that, in some sense, comprises all conditions which are not explicitly defined in the model. But then, analogously to defined unobserved variables, there is no immediate reason to use a stochastic variable $\dot{Z}$; instead one should begin with a deterministic variable

$\ddot{Z}$ and consequently with a model of the form $\ddot{X} \longrightarrow \ddot{Y} \longleftarrow \ddot{Z}$. Only the intention to use a reduced model instead might then provide a reason for substituting $\ddot{Z}$ by a stochastic variable $\dot{Z}$. However, the reduced model is then of the form (4.4), and $\dot{Z}$ no longer occurs.

This consideration also allows clarification of the question whether one can assume that the residuals (i.e. the stochastic exogenous variables) of a pseudo-indeterministic model are stochastically independent of its explicitly defined exogenous variables; in the example, whether $\dot{Z}$ is stochastically independent of $\ddot{X}$.[25] In a sense, the question has no point because it vanishes as soon as one considers a stochastically equivalent reduced model.[26] Of course, beginning with model (4.5), it would be possible to assume that $\dot{Z}$ might be stochastically dependent on $\ddot{X}$. But then results a new model that has the following form:

$$\begin{tikzcd} \ddot{X} \arrow[d] \arrow[dr] & \\ \dot{Z} \arrow[r] & \dot{Y} \end{tikzcd} \tag{4.10}$$

$\dot{Z}$ then becomes an endogenous variable, and the model is no longer pseudo-indeterministic. In fact, the resulting model only has a single deterministic exogenous variable. Assuming that residuals might be stochastically dependent on explicitly defined exogenous variables therefore implies the presupposition that a pseudo-indeterministic model would not be adequate.

5 Functional causality

5.1 Functional causes and conditions
1. Deterministic models
2. Definitions for stochastic models
3. True and spurious causes?
4. Singular and generic causal statements
5. Retrospective and prospective questions
6. Dynamic and comparative causes

5.2 Ambiguous references to individuals
1. Holland's definition of causal effects
2. Two kinds of questions
3. Different homogeneity assumptions
4. Generic effects for individuals

5.3 Isolating functional causes
1. Deterministic covariates
2. Stochastic covariates
3. Interaction and distribution-dependence
4. Exogenous and endogenous causes
5. Direct and indirect effects

This chapter discusses notions of "causal relationships" in the framework of functional models. The leading idea is to distinguish between values of variables that can be viewed as *conditions* of values of other variables and changes of values of variables that can be viewed as *causes* of changes of values of other variables. Since the definitions relate to functional models this approach is termed *functional causality*. It provides a formal framework within which further distinctions (in particular between comparative and dynamic conceptions) can be discussed.

The chapter contains three sections. The first section discusses definitions of functional causality for deterministic and stochastic models. It is stressed that the definitions relate to models and not immediately to the empirical world, and some implications and restrictions due to this fact are considered. The second section deals with an approach, currently widespread in the statistical literature, that tries to define a notion of causality in terms of counterfactual (statistical or stochastic) variables. The third section is concerned with the idea that it should be possible to isolate causes and attribute to them specific effects. It is argued that this is not always possible, resulting in some limitations for the applicability of causal ideas.

5.1 Functional causes and conditions

1. Deterministic models

Let $\mathcal{M} = (\mathcal{V}, \mathcal{F})$ denote a deterministic functional model and $\ddot{X}$ and $\ddot{Y}$ two variables in $\mathcal{V}$. The goal is to explicate what is meant when saying that a change $\Delta(x', x'')$ in variable $\ddot{X}$ is a *functional cause* of a change $\Delta(y', y'')$ in the variable $\ddot{Y}$.[1]

A *covariate context* of a variable $\ddot{X}$ having the value $\ddot{X} = x$ with respect to a variable $\ddot{Y}$ (which is functionally dependent on $\ddot{X}$) consists of a variable $\ddot{Z} \in \mathcal{V}$ such that, if $\ddot{Z}$ takes the value z, the value of $\ddot{Y}$ can be uniquely determined from the knowledge of $\ddot{X} = x$ and $\ddot{Z} = z$.[2] Of course, it might be possible to derive a unique value of $\ddot{Y}$ already from a knowledge of $\ddot{X} = x$ alone. With the help of this notion a deterministic version of functional causality can be defined:

> A change $\Delta(x', x'')$ in the variable $\ddot{X}$ in the covariate context $\ddot{Z} = z$ is a *functional cause* of a change $\Delta(y', y'')$ in the variable $\ddot{Y}$ if $\ddot{Y}$ is functionally dependent on $\ddot{X}$, $\ddot{Y} = y'$ can be derived from $\ddot{X} = x'$ and $\ddot{Z} = z$, and $\ddot{Y} = y''$ can be derived from $\ddot{X} = x''$ and $\ddot{Z} = z$. If these derivations do not depend on a covariate context, $\Delta(x', x'')$ is called a *context-independent functional cause* of $\Delta(y', y'')$.

Correspondingly, $\Delta(y', y'')$ is called a *functional effect* of its (context-dependent or -independent) functional cause. The following remarks may clarify the definitions.

(a) Statements about functional causes and effects always relate to a functional model. These notions can therefore not be used for statements which directly refer to the empirical world. This is also important for an understanding of formulations which refer to changes of values of a variable. Such changes as, for example, a change $\Delta(x', x'')$ in a variable $\ddot{X}$, result from assumptions, or some other kind of mental operation, which a person performs in the context of a model.

(b) Functional causality connects *changes* of the values of variables.[3] Following the suggested definition, neither variables nor values of variables can be viewed as causes.[4] It is often possible, however, to think of values of a variable as being *functional conditions* of values of some other variable.[5] Referring to the example discussed in § 1 of Section 4.1, one can reasonably say that the switch's being closed ($\ddot{X} = 1$) is a functional condition for the bulb's giving light ($\ddot{Y} = 1$). However, only a change like closing the switch ($\Delta(0, 1)$ in the variable $\ddot{X}$) can possibly (in the covariate context $\ddot{Z} = 1$, that is, if the battery provides power) be called a functional cause of another change ($\Delta(0, 1)$ in the variable $\ddot{Y}$).[6]

(c) While the idea of restricting functional causes to changes and to distinguish them from conditions is of fundamental importance, it is unfortunately also a source of potential confusions. This is due to the fact that a substantial notion of change cannot readily be formulated with an abstract functional model. Instead, it is often only possible to compare different functional conditions.[7]

(d) Functional causality is in most cases context-dependent: Whether, and how, a change $\Delta(x', x'')$ is a functional cause of a change $\Delta(y', y'')$ most often also depends on values of further variables. For example, closing the switch is only a cause for the bulb's giving light if the battery provides power. This is in accordance with the fact, stressed in Section 4.1 (§ 3), that the dependencies in a functional model are given by functions which often have several arguments. The arrows used in the graphical illustrations of these functions can therefore not, in general, be associated with specific causal relations.

(e) In order for a change $\Delta(x', x'')$ in a variable $\ddot{X}$ to be a functional cause of a change in the values of another variable $\ddot{Y}$, the change $\Delta(x', x'')$ must be possible without a simultaneous change in the values of other variables on which $\ddot{Y}$ depends. This is not always true if $\ddot{X}$ is an endogenous variable or if there are constraints for the covariation of $\ddot{X}$ and some other variables. Further restrictions can result from context-dependencies because it must be possible, then, to have the same covariate context $\ddot{Z} = z$ both with $\ddot{X} = x'$ and $\ddot{X} = x''$. Restrictions result if $\ddot{Z}$ depends on $\ddot{X}$; there is then no effect which can be uniquely attributed to the change $\Delta(x', x'')$. (Such problems will be further discussed in Section 5.3.)

(f) Since functional models do not require a temporal interpretation of their functions the definition proposed above does not require that a cause precedes its effects in time. It is quite possible, however, to add this demand when dealing with dynamic models that can be viewed as process frames based on a time axis.

The idea to distinguish between causes and conditions allows for a better understanding of the often made statement that events result, in most cases, from a multitude of causes.[8] It is, of course, always possible to think of any number of conditions on which the occurrence of an event depends; however, the intention of causal statements normally is to learn about one, or a few, causes which *under given conditions* led to the event. It is mainly for this reason that the notion of cause requires that context conditions can be held constant.

2. Definitions for stochastic models

The definitions of functional causality in stochastic models must take into account the type of function. If two stochastic variables are connected by a deterministic function, for example $\dot{X} \longrightarrow \dot{Y}$ (where further variables may be added to provide a covariate context), one can directly apply the definition of § 1. A different definition is required if the possible effects relate to an endogenous stochastic variable $\dot{Y}$ whose dependency on other variables is specified by a stochastic function.

To begin with, consider a simple model having the form $\ddot{X} \longrightarrow\!\!\!\!\rightarrow \dot{Y}$, for example (as in § 2 of Section 4.2), a toaster where $\ddot{X}$ and $\dot{Y}$ record, respectively, the position of a lever and the toasting duration. A change $\Delta(x', x'')$ is then associated, not with a single change in $\dot{Y}$, but with two conditional distributions: $\Pr[\dot{Y}|\ddot{X} = x']$ and $\Pr[\dot{Y}|\ddot{X} = x'']$. There are, therefore, different possibilities to define causal effects of a change $\Delta(x', x'')$. Because there is no context-dependence in this example, it would be possible to use a difference of expected values:

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

However, depending on the application, different definitions can be useful and therefore the general notation $\Delta^s(\dot{Y}; x', x'')$ will be used to denote a *stochastic effect* which is in some way defined with respect to a stochastic variable $\dot{Y}$ and a change $\Delta(x', x'')$ of another variable.

Using this general notation, and adopting accordingly the meaning of covariate context from § 1, functional causes of stochastic effects can be defined in the following way (it is assumed that the variables used in the definitions are defined for some stochastic model $\mathcal{M} = (\mathcal{V}, \mathcal{F})$):

> A change $\Delta(x', x'')$ in the variable $\ddot{X}$ in the covariate context $\ddot{Z} = z$ is a *functional cause* of a stochastic effect $\Delta^s(\dot{Y}; x', x'')$ if a directed path leads from $\ddot{X}$ to $\dot{Y}$, and the stochastic effect can be determined from the conditional distributions $\Pr[\dot{Y}|\ddot{X}=x', \ddot{Z}=z]$ and $\Pr[\dot{Y}|\ddot{X}=x'', \ddot{Z}=z]$. $\Delta(x', x'')$ is called a *context-independent functional cause* of the stochastic effect if its derivation only requires the conditional distributions $\Pr[\dot{Y}|\ddot{X}=x']$ and $\Pr[\dot{Y}|\ddot{X}=x'']$. (Both definitions can be used accordingly if $\ddot{X}$ and/or $\ddot{Z}$ are stochastic.)

The comments (a)–(f) of § 1 apply accordingly to stochastic effects. It should be stressed, again, that functional causes are defined as changes $\Delta(x', x'')$.[9] It is also important that stochastic effects are most often context-dependent. It is quite possible, for example, that a change $\Delta(x', x'')$ in one covariate context increases the probability of an event while it decreases the probability in another covariate context. Context-independent definitions of probabilistic causality, while often attempted, are thus difficult to justify.[10] In particular, it seems not reasonable to require by definition that a cause always increases (in all possible contexts) the probability of an event.[11]

3. *True and spurious causes?*

In discussions of probabilistic causality it is often deemed necessary to distinguish between true and spurious causes. The idea is that a supposed cause of some effect is spurious if it becomes superfluous for an explanation of the effect when further conditions are taken into account.[12] However, the distinction cannot reasonably be applied if one uses the definition of functional causes proposed in § 2. Given the necessary context, a functional cause is always a "true" cause. Of course, whether a change $\Delta(x', x'')$ is a functional cause of some stochastic effect $\Delta^s(\dot{Y}; x', x'')$ also depends on the assumed model. However, this is actually a truism since functional causal statements always relate to a model.

4. Singular and generic causal statements

Singular causal statements refer to specific situations which can be identified in the empirical world; for example: The stone's impact on the window-pane was the cause of its break. Causal statements of this kind presuppose the reference to a situation that allows the speaking of specific events. Generic causal statements, on the other hand, refer to event types (or, more abstract, to changes in the values of variables); for example: a window-pane probably breaks if hit by a stone. This is a general statement that does not refer to any specific events occurring in an empirically identifiable situation. Obviously, the definitions of functional causes and effects proposed above lead to generic causal statements.

Generic causal statements do not directly refer to (aspects of) the empirical world, but to models. On the other hand, singular causal statements refer to real world situations. A singular causal statement presupposes a real world context in which not just the events referred to in the causal statement can be empirically identified, but any number of additional aspects of the situation can also be identified if that would be necessary in order to explicate the statement.

The situation is quite different with generic causal statements. Since they do not relate to any specific situation in the empirical world, there is no chance to add (empirical) references if that would be required for an explication. Instead, the context on which the statement's meaning and validity depend must be provided in a general way, that is, in the framework of a model, as conditions that can be specified by values of variables. Since only very few conditions can be explicitly stated in a model, the meaning as well as the validity of generic causal statements always remain confined to a model and cannot without qualifications be applied to the empirical world. In particular, it is not possible to derive (in any valid sense) singular from generic causal statements. This does not, of course, exclude the possibility of using generic causal statements as arguments in the reasoning about singular causal questions.

5. Retrospective and prospective questions

A further distinction concerns the kind of question to which causal statements answer. Retrospective questions ask how (by which causes) some fact came into being. The question presupposes that the fact, to be explained, already happened and can be empirically identified; the question only concerns whether, and how, a causal explanation can be given. Prospective questions, on the other hand, ask for the possible effects of some cause. Questions of this kind presuppose, hypothetically or as a matter of fact, the realization of some cause and ask what effects will probably result.

It has been suggested that only prospective causal questions can be answered unambiguously.[13] However, as will be shown in Section 5.3, prospective questions do not always permit an unambiguous answer. On the other hand, retrospective questions can often be answered within a presupposed model. For example, given the model mentioned in § 1 of Section 4.1, if one knows that the battery provides power, the only possible cause for the bulb's ceasing to provide light is that the switch has been turned off. Of course, in many cases one can imagine several different causes which might have brought about a fact (to be explained) and it is then impossible to give an unambiguous explanation.

6. Dynamic and comparative causes

The definitions of functional causes given at the beginning of this section provide a formal framework compatible with substantially different notions of cause. In particular, the framework covers both dynamic and comparative notions of cause. To illustrate the distinction, consider a change $\Delta(x', x'')$ defined for some variable $\ddot{X}$.

- $\Delta(x', x'')$ can be called a *dynamic cause* (of some effect) if one can imagine an event during which the value of $\ddot{X}$ changes from x' to x''. For example, a switch is closed, a person becomes unemployed, a patient takes a drug.
- In contrast, $\Delta(x', x'')$ will be called a *comparative cause* if the values x' and x'' relate to two different situations (or objects), say $\sigma(x')$ and $\sigma(x'')$, and one is unwilling or unable to think of an event, or process, by which $\sigma(x'')$ developed from $\sigma(x')$.

Statistical social research is predominantly concerned with comparative causes. The standard approach consists in comparing people, or situations, which can be characterized by different properties. One compares, for example, the salaries of people having different levels of education; and this leads to the conclusion that differences in the educational level are a (possible) cause of differences in salaries. This would be an example of a comparative cause and should be distinguished from a dynamic cause which, in this example, would consist in a change of a person's educational level.

5.2 Ambiguous references to individuals

The definitions of functional causality proposed in the previous section are based on functional models. These models are formulated in terms of generic variables and therefore do not directly apply to identifiable individuals (objects or situations). The same is true of the causal statements which can be derived from these models. This also distinguishes our approach from the potential outcomes approach that was proposed, in particular, by D. B. Rubin, P. W. Holland, and

P. R. Rosenbaum.[14] The present section tries to show that some of the crucial difficulties of the potential outcomes approach mainly result from its attempt to directly explicate causal statements by referring to (empirically) identifiable individuals.[15]

1. Holland's definition of causal effects

The following discussion refers to an explication of the potential outcomes approach developed by P. W. Holland. The starting point is a finite collection of individuals that will be denoted by $\Omega = \{\omega_1, \ldots, \omega_n\}$. A property space $\tilde{\mathcal{X}} = \{\tilde{x}_1, \ldots, \tilde{x}_m\}$ serves to define possible causes. The elements of $\tilde{\mathcal{X}}$ denote different treatments which can be applied to the individuals in Ω. A selection of treatments is then given by a statistical variable

$$X : \Omega \longrightarrow \tilde{\mathcal{X}}$$

with $X(\omega)$ being the treatment applied to ω. A further property space, $\tilde{\mathcal{Y}}$, is used to record possible effects. Holland's basic idea is to assume, for each value $\tilde{x} \in \tilde{\mathcal{X}}$, the existence of a statistical variable

$$Y_{\tilde{x}} : \Omega \longrightarrow \tilde{\mathcal{Y}}$$

having the following interpretation: $Y_{\tilde{x}}(\omega)$ is the outcome which would result if the treatment $\tilde{x}$ would have been applied to ω. Holland assumes a deterministic relationship, but this is not crucial and stochastic formulations would also be possible. Of crucial importance is, however, that Holland assumes that for each individual $\omega \in \Omega$ all possible effects $Y_{\tilde{x}_1}(\omega), \ldots, Y_{\tilde{x}_m}(\omega)$ exist in some sense simultaneously and can be used in subsequent theoretical derivations. Due possibly to the counterfactual formulations used by Holland, this seems to be an obscure assumption.[16] It is quite possible, however, to begin with a more abstract formulation:

Assumption: For every individual $\omega \in \Omega$ exists a function: $f_\omega : \tilde{\mathcal{X}} \longrightarrow \tilde{\mathcal{Y}}$ (5.1)

Given such functions, they immediately provide values of the statistical variables supposed by Holland: $Y_{\tilde{x}}(\omega) = f_\omega(\tilde{x})$. One has to note, of course, that (5.1) is not an empirical assumption but actually implies a separate functional model for each individual in Ω.[17]

As an important implication, the assumption (5.1) allows to define *individual* causal effects:

$$\Delta^i(Y; x', x''; \omega) := f_\omega(x'') - f_\omega(x') = Y_{x''}(\omega) - Y_{x'}(\omega) \tag{5.2}$$

where x' and x'' are any values in $\tilde{\mathcal{X}}$. And this definition finally creates Holland's (1986: 947) "fundamental problem of causal inference": that values of such causal effects can never be observed.

2. Two kinds of questions

In order to understand Holland's "fundamental problem," one should distinguish two kinds of questions. Questions of the first kind will be called *concrete modal questions* because they ask for possible consequences of an event that can be identified in a concrete situation; for example: How will the disease of a patient probably develop if a specific therapy is applied?[18]

It is obvious that questions of this kind compare two (or more) possibilities and that at best one realization can be observed. It is also obvious, however, that the problem does not result from observational limitations. The problem rather concerns justifications of conditional statements which correspond to the different possibilities. While the necessary knowledge can only result from observations, these observations cannot result from the situation referred to by the modal question. Instead, the necessary empirical knowledge can only result from observations of comparable situations which took place in the past.

This then motivates the development of models as means to make observations of previous situations useful for considerations of modal questions. But then the kind of question changes as well. Instead of concrete questions which refer to a concrete situation, one has to consider *generic modal questions* which refer to generic situations that cover all situations of a specific type (including the previous situations that generated the data).

Following this line of reasoning, Holland's "fundamental problem" actually becomes irrelevant, as can be seen by a comparison of the following model frameworks.

$$\text{(a)}\quad \begin{array}{l} \ddot{X} \searrow \\ \qquad\quad \ddot{Y} \\ \ddot{I} \nearrow \end{array} \qquad\qquad \text{(b)}\quad \begin{array}{l} \ddot{X} \searrow \\ \qquad\quad \dot{Y} \\ \ddot{Z} \nearrow \end{array} \tag{5.3}$$

Holland's approach corresponds to (a). $\ddot{X}$ records treatments defined in $\tilde{\mathcal{X}}$, and $\ddot{Y}$ records outcomes which can be observed afterwards. In addition, there is a variable $\ddot{I}$ that serves to identify individual members of the presupposed collection Ω.[19] But as a consequence of using this variable $\ddot{I}$, the model can only be used to represent data which might be available for the individuals in Ω; it does not allow the formulation of any generic question. Furthermore, it also cannot be used for a concrete modal question that refers to some individual ω^*, simply because values of $\ddot{X}$ are already fixed for all individuals in Ω, and ω^* cannot, therefore, be a member of Ω.

In contrast, the approach in (5.3b) corresponds to a generic modal question. It is based on a statistical view that distinguishes between individuals only insofar as they have distinct values for some variables (having ranges defined without reference to individual names).[20] In the example a variable $\ddot{Z}$ having some range $\tilde{\mathcal{Z}}$ is used. The model no longer relates to the members of some specific collection Ω, but in an abstract way to individuals which can be characterized by values of $\ddot{Z}$. The model simply does not distinguish between individuals having the same value of $\ddot{Z}$. Since it is quite possible that individuals with identical values of $\ddot{X}$

and $\ddot{Z}$ generate different values in $\tilde{\mathcal{Y}}$, the model uses a stochastic variable $\dot{Y}$ that is stochastically dependent on $\ddot{X}$ and $\ddot{Z}$.

3. Different homogeneity assumptions

A further question concerns the assumptions needed to justify estimation of (individual) causal effects with statistical data. In Holland's approach, the crucial assumption is that the functions f_ω are identical for all individuals in Ω (or in subgroups defined by generic variables).[21] This assumption, however, conflicts with a statistical view since it implies that all individuals which can be characterized by identical values of some variables therefore behave identically.

In contrast, the statistical view posits that it is sometimes, for some kinds of questions, reasonable to consider only a few differences between individuals and abstract from any other differences which certainly exist as well. To this idea corresponds the model (5.3b). The model obviously does not require the assumption that the functions f_ω are identical for all individuals having the same value of $\ddot{Z}$. In fact, it is not even necessary to assume that these functions actually do exist. Instead of (5.1), the model requires a statistical homogeneity assumption: that there exists, for each value $\ddot{Z} = z$, a stochastic function

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

Thus $\ddot{Z}$ figures as a covariate context for functional causal relationships between $\ddot{X}$ and $\dot{Y}$.

4. Generic effects for individuals

It is noteworthy that the model (5.3b) does not permit definition of variables of the form $Y_{\tilde{x}}$ (see § 1). Instead, one can consider variables

$$Y^*_{\tilde{x}} : \Omega \longrightarrow \mathbf{R}$$

defined by $Y^*_{\tilde{x}}(\omega) := \mathrm{E}(\dot{Y} \,|\, \ddot{X} = \tilde{x}, \ddot{Z} = Z(\omega))$.[22] Generic causal effects for individuals can then be defined as

$$\Delta^*(\dot{Y}; x', x''; \omega) := Y^*_{x''}(\omega) - Y^*_{x'}(\omega) \tag{5.4}$$

In contrast to the effects defined in (5.2), these are now generic effects which are not defined for identifiable individuals but for equivalence classes induced by $\ddot{Z}$. Of course, referring to a specified collection Ω and using assumption (5.1), these generic effects can be expressed as mean values of the individual effects defined in (5.2):

$$\Delta^*(\dot{Y}; x', x''; \omega) = \frac{1}{|\{\omega' | \ddot{Z}(\omega') = \ddot{Z}(\omega)\}|} \sum_{\omega' \in \{\omega' | \ddot{Z}(\omega') = \ddot{Z}(\omega)\}} \Delta^i(Y; x', x''; \omega')$$

However, this interpretation rests on the assumption (5.1) which is not required in order to interpret the generic effects defined in (5.4). And since it requires the reference to a specified collection of individuals, this interpretation is compatible neither with concrete nor with generic causal questions.

5.3 Isolating functional causes

Notions of (functional) causality usually imply the idea that one can isolate causes and attribute to them specific effects.[23] The definitions proposed in Section 5.1 adhere to this idea by explicitly referring to a possible covariate context. Furthermore, in order to interpret a change $\Delta(x', x'')$ as a cause (of some effect), it must be possible to keep the covariate context constant.[24] However, as will be discussed in the present section, this is not always possible, in particular when trying to associate causes with endogenous variables of a model.

1. Deterministic covariates

Consider a simple model with a single endogenous variable, $\dot{Y}$, having a distribution which depends on values of deterministic variables $\ddot{X}$ and $\ddot{Z}$. The dependence is specified by a stochastic function

$$(x, z) \longrightarrow \Pr[\dot{Y} | \ddot{X} = x, \ddot{Z} = z] \tag{5.5}$$

The stochastic effect of a change $\Delta(x', x'')$ in the variable $\ddot{X}$ can then be defined, for example, by

$$\Delta^s(\dot{Y}; x', x''; z) := \mathrm{E}(\dot{Y} | \ddot{X} = x'', \ddot{Z} = z) - \mathrm{E}(\dot{Y} | \ddot{X} = x', \ddot{Z} = z) \tag{5.6}$$

and the formulation directly shows how the effect depends on the covariate context $\ddot{Z} = z$. The definition is based on the assumption that a change $\Delta(x', x'')$ in the variable $\ddot{X}$ is compatible with a constant covariate context $\ddot{Z} = z$. This is required in order to think of $\Delta(x', x'')$ as a specific cause that can be isolated. However, one can easily imagine models in which it is impossible to hold constant a covariate context. This may result, for example, from constraints that restrict the possibilities to vary values of $\ddot{X}$ independently of values of $\ddot{Z}$.

Another situation of interest is when $\ddot{Z}$ depends on values of $\ddot{X}$. Assuming for the moment a deterministic relationship, the model then takes the following form:

$$\begin{tikzcd} \ddot{X} \arrow[dd] \arrow[dr, ->>] & \\ & \dot{Y} \\ \ddot{Z} \arrow[ur, ->>] & \end{tikzcd} \tag{5.7}$$

In addition to the function (5.5) there is a deterministic function $z = g(x)$. Consequently, the distribution of $\dot{Y}$ depends only on values of $\ddot{X}$:

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

Even though in this case a change $\Delta(x', x'')$ is not compatible with a constant covariate context, it can nevertheless be interpreted as the functional cause of a stochastic effect

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

The reason is that $\ddot{Z}$ deterministically depends on $\ddot{X}$ so that $\ddot{Z}$ does not figure as a (in some sense autonomous) covariate context. This implies in turn that changes in the values of $\ddot{Z}$ cannot be interpreted as possible causes of stochastic effects defined with respect to $\dot{Y}$.

2. Stochastic covariates

Somewhat different considerations are required if the covariate context is given by a stochastic variable $\dot{Z}$. To begin with, consider a model

$$\begin{tikzcd} \ddot{X} \arrow[dr, two heads] & \\ & \dot{Y} \\ \dot{Z} \arrow[ur, two heads] & \end{tikzcd} \tag{5.8}$$

where $\dot{Z}$, with distribution $\Pr[\dot{Z}]$, is stochastically independent of $\ddot{X}$. The mean effect of $\Delta(x', x'')$ may be defined in the following way:

$$\begin{aligned} &\Delta^s(\dot{Y}; x', x''; \dot{Z}) := \\ &\quad \textstyle\sum_z (\mathrm{E}(\dot{Y}|\ddot{X} = x'', \dot{Z} = z) - \mathrm{E}(\dot{Y}|\ddot{X} = x', \dot{Z} = z)) \Pr(\dot{Z} = z) \end{aligned} \tag{5.9}$$

where the mean is calculated with respect to the distribution of $\dot{Z}$. Obviously, this mean effect equals the effect of $\Delta(x', x'')$ in the reduced model $\ddot{X} \twoheadrightarrow \dot{Y}$ which can be derived from (5.8).

3. Interaction and distribution-dependence

It is quite possible that the stochastic effect defined in (5.9) also depends on the distribution of $\dot{Z}$. This is the case if $\ddot{X}$ and $\dot{Z}$ are *interactive conditions* (for the distribution of $\dot{Y}$) defined as follows:

> $\ddot{X}$ and $\ddot{Z}$ are *interactive conditions* for $\dot{Y}$ if the stochastic effect $\Delta^s(\dot{Y}; x', x''; z)$ defined in (5.6) depends on z, that is, if there exist at least two values z' and z'' such that $\Delta^s(\dot{Y}; x', x''; z') \neq \Delta^s(\dot{Y}; x', x''; z'')$.
>
> If the covariate context is stochastic, the same definition applies with $\dot{Z}$ replacing $\ddot{Z}$ in (5.6). A further definition refers to the stochastic effect defined in (5.9). It will be said that this effect is *distribution-dependent* if it depends on the distribution of $\dot{Z}$.[25]

Linear regression functions can serve to illustrate the definitions. Given a regression function

$$\mathrm{E}(\dot{Y}|\ddot{X}=x,\dot{Z}=z)=\alpha+x\beta_x+z\beta_z$$

$\ddot{X}$ and $\dot{Z}$ are not interactive conditions; the effect of a change of the values of $\ddot{X}$ does not depend on $\dot{Z}$. If one extends the model by adding an interaction term:

$$\mathrm{E}(\dot{Y}|\ddot{X}=x,\dot{Z}=z)=\alpha+x\beta_x+z\beta_z+xz\beta_{xz}$$

$\ddot{X}$ and $\dot{Z}$ become interactive conditions and the stochastic effect will be distribution-dependent:

$$\Delta^s(\dot{Y};x',x'';\dot{Z})=(x''-x')(\beta_x+\beta_{xz}\,\mathrm{E}(\dot{Z}))$$

As a simple illustration, let $\dot{Z}$ record the type (1 or 2) of the school visited by a child, $\ddot{X}$ records the parents' educational level (0 low, 1 high), and $\dot{Y}$ records the child's educational success (0 not successful, 1 successful). The following conditional expectations will be assumed:

x	z	$\mathrm{E}(\dot{Y}\mid\ddot{X}=x,\dot{Z}=z)$
0	1	0.8
0	2	0.6
1	1	0.8
1	2	0.9

(5.10)

In this example, the parents' educational level is of relevance for the child's educational success in school type 2, but not in school type 1. $\ddot{X}$ and $\dot{Z}$ are then interactive conditions: $\Delta^s(\dot{Y};0,1;1)=0$ and $\Delta^s(\dot{Y};0,1;2)=0.3$; and the effect is also distribution-dependent:

$$\Delta^s(\dot{Y};0,1;\dot{Z})=\Delta^s(\dot{Y};0,1;2)\,\mathrm{Pr}(\dot{Z}=2)$$

The probability of the child's success depends on the probability of visiting one or the other school type. The example nevertheless assumes that the distribution of $\dot{Z}$ does not depend on values of $\ddot{X}$. As mentioned in § 2, this is required in order to think of $\Delta(x',x'')$ as a cause which can be isolated. Distribution-dependence can then be viewed as a form of dependency on a covariate context (which parallels the dependency on values of deterministic covariates).

4. Exogenous and endogenous causes

A possible cause can relate to an exogenous or to an endogenous variable of a model. The change results in both cases from an assumption: that the value of a variable changes from x' to x''; and this assumption only requires that, in the given model, both values must be possible. Of course, if the change relates to an endogenous variable, one has to take into account that its values depend on values or distributions of other variables.

If the endogenous variable is deterministic one can consider, for example, a model

$$\ddot{X} \longrightarrow \ddot{Z} \longrightarrow\!\!\!\!\!\rightarrow \dot{Y} \tag{5.11}$$

where $\ddot{X}$ and $\ddot{Z}$ are connected by a deterministic function $z = g(x)$. It seems quite possible, then, to think of a change $\Delta(z', z'')$ in the endogenous variable $\ddot{Z}$ as being the cause of some stochastic effect defined with respect to $\dot{Y}$. The reason is that, if z' and z'' are possible values of $\ddot{Z}$, then also there exist values x' and x'' of $\ddot{X}$ such that $z' = g(x')$ and $z'' = g(x'')$. Note that this is different from the example considered in § 1 where $\dot{Y}$ depends also directly on $\ddot{X}$.[26]

Now consider a model with a stochastic endogenous variable $\dot{Z}$:

$$\ddot{X} \longrightarrow\!\!\!\!\!\rightarrow \dot{Z} \longrightarrow\!\!\!\!\!\rightarrow \dot{Y} \tag{5.12}$$

In the same way as was done before one can interpret changes $\Delta(z', z'')$ as causes of stochastic effects defined with respect to $\dot{Y}$. There is, however, a remarkable difference. The model (5.11) allows the creation of a cause $\Delta(z', z'')$ by some appropriate manipulation of the exogenous variable $\ddot{X}$; but this is not possible in the model (5.12).

5. *Direct and indirect effects*

Finally, consider a situation where $\dot{Y}$ also directly depends on $\ddot{X}$. The model has then the form

$$\begin{tikzcd}
\ddot{X} \arrow[d, two heads] \arrow[dr, two heads] & \\
 & \dot{Y} \\
\dot{Z} \arrow[ur, two heads] &
\end{tikzcd} \tag{5.13}$$

A modification of the example considered in § 3 can serve as an illustration. $\dot{Z}$ (with values 1 or 2) records the school type, $\ddot{X}$ (with values 0 or 1) records the parents' educational level, and $\dot{Y}$ (with values 0 or 1) records the child's educational success. In the modified example the child's school type depends on the parents' educational level. The following values will be assumed for the illustration:

x	z	$\mathrm{E}(\dot{Y} \mid \ddot{X} = x, \dot{Z} = z)$
0	1	0.5
0	2	0.7
1	1	0.8
1	2	0.9

x	$\Pr(\dot{Z} = 2 \mid \ddot{X} = x)$
0	0.4
1	0.8

(5.14)

Comparing this model with (5.12), one observes that effects of a change $\Delta(z', z'')$ are now context-dependent. For example, corresponding to a change $\Delta(1, 2)$ one finds:

$$\mathrm{E}(\dot{Y}|\dot{Z}=2, \ddot{X}=0) - \mathrm{E}(\dot{Y}|\dot{Z}=1, \ddot{X}=0) = 0.7 - 0.5 = 0.2$$

$$\mathrm{E}(\dot{Y}|\dot{Z}=2, \ddot{X}=1) - \mathrm{E}(\dot{Y}|\dot{Z}=1, \ddot{X}=1) = 0.9 - 0.8 = 0.1$$

The model (5.13) also suggests an attempt to distinguish direct, indirect, and total effects of a change $\Delta(x', x'')$ in the variable $\ddot{X}$. However, that will not always be possible. Consider, for example, the total effect of a change $\Delta(0, 1)$:[27]

$$\mathrm{E}(\dot{Y}|\ddot{X}=1) - \mathrm{E}(\dot{Y}|\ddot{X}=0) = 0.88 - 0.58 = 0.3$$

If one considers $\dot{Z}$ as providing a covariate context and assumes that values of $\dot{Z}$ can be fixed, one can also calculate direct effects; in the example:

$$\mathrm{E}(\dot{Y}|\ddot{X}=1, \dot{Z}=1) - \mathrm{E}(\dot{Y}|\ddot{X}=0, \dot{Z}=1) = 0.8 - 0.5 = 0.3$$

$$\mathrm{E}(\dot{Y}|\ddot{X}=1, \dot{Z}=2) - \mathrm{E}(\dot{Y}|\ddot{X}=0, \dot{Z}=2) = 0.9 - 0.7 = 0.2$$

These effects are obviously context-dependent and it is not possible, therefore, to think of just one direct (context-independent) effect of the parents' educational level on the child's educational success. As a consequence, it is also not possible to calculate an indirect effect that results from the different probabilities for the selection of school types. The example shows that total effects cannot always clearly be separated into direct and indirect effects.

6 Models and statistical data

6.1 Functional models and data
1. Functional models and data models
2. Data generation with functional models
3. Functional models for statistical data
4. Indeterminate mean effects
5. Individuals and populations

6.2 Experimental and observational data
1. Randomized experiments
2. What can be achieved with randomization?
3. The standard argument for randomization
4. Limitations of using mean effects
5. Arguments based on potential outcomes
6. Causal and pseudo-descriptive questions
7. Observational data

6.3 Interventions and reference problems
1. Causal knowledge and interventions
2. Causes as parts of processes
3. Different kinds of interventions
4. Assumptions about people's behavior
5. Relevance of the substantial actors
6. Distortions due to self selection?
7. Selection problems from expectations?
8. Fictitious stochastic anticipations
9. Presuppositions about invariances

Functional models are theoretical constructs. They do not entail anything about the facts of the empirical world, nor are they rendered false by any facts. But reference to the empirical world is indispensable if a model is to be used for (retrospective) explanations or (prospective) predictions; and being interested in quantitative statements the model's deterministic or stochastic functions need numerical specification. There are thus often reasons to use statistical data to calculate estimates of the functions of a model.

This chapter discusses the relationship between functional models and statistical data. The discussion is conceptual; techniques for the estimation of model parameters will not be treated. The first section examines the dichotomy between functional models and statistical data. The second section considers the question whether, and how, statistical data can be used to learn about causal relationships defined within functional models. It is shown that the conceptual framework of randomized experiments is of only limited use for social science applications. The third section argues that several reference problems relevant for understanding empirical claims to be connected with functional models should be distinguished.

6.1 Functional models and data

1. Functional models and data models

Considered as analytical models, functional models express relationships between modal variables and serve to consider modal questions. In contrast, *data models* (as the expression is used here) relate to statistical data representing facts realized in the past. The statistical data are represented by a (multi-dimensional) statistical variable, say S, which is defined for a reference set Ω.

- A *statistical data model* describes the statistical distribution $\mathrm{P}[S]$, or some aspects of this distribution (e.g. a regression function that can be derived from $\mathrm{P}[S]$).
- A *stochastic data model* invents and describes a random generator such that one can imagine the data, S, to be a realization of the generator.[1] The random generator may be seen as a model for a data generating (and sometimes also a substantial) process that might have produced the data.[2]

Stochastic data models can be viewed as a particular way to connect functional models with statistical data. Starting from a functional model one creates a stochastic model for the data. In the simplest case (assumed for the illustrations in this chapter) one can then directly use the (conditional) statistical distributions derived from the data as estimates of correspondingly defined (conditional) probability distributions of the functional model.

2. Data generation with functional models

Given a functional model $\mathcal{M} = (\mathcal{V}, \mathcal{F})$, it can be used to generate data. There are four steps.

(1) The first step is the definition of a reference set, Ω, whose elements can be used as units (objects or situations) to be associated with valuations of the model's variables.
(2) Corresponding to the variables in $\mathcal{V}$, one then defines statistical variables for Ω. If, for example, $\mathcal{V}$ contains variables $\ddot{X}$, $\ddot{Z}$, and $\dot{Y}$ with domains $\tilde{\mathcal{X}}$, $\tilde{\mathcal{Z}}$, and

$\tilde{\mathcal{Y}}$, respectively, one defines a three-dimensional statistical variable

$$(X, Z, Y) : \Omega \longrightarrow \tilde{\mathcal{X}} \times \tilde{\mathcal{Z}} \times \tilde{\mathcal{Y}}$$

such that $(X, Z, Y)(\omega)$ records the values of $\ddot{X}$, $\ddot{Z}$, and $\dot{Y}$ that will be realized for the unit ω.

(3) One then creates for each $\omega \in \Omega$ values for the statistical variables that correspond to the model's exogenous variables. If the variables are stochastic this is done with appropriate random generators, otherwise values can be created arbitrarily (observing, of course, any constraints which might be defined in the model).

(4) Finally, beginning with the values created in step (3), one uses the functions given in $\mathcal{F}$ in order to derive (recursively) values of the remaining variables. Again, if the functions are stochastic, one uses appropriately defined random generators.

It is remarkable that in step (3) values of the deterministic exogenous variables can be generated arbitrarily. It follows in particular that one can assume arbitrary distributions for the correspondingly defined statistical variables and, in turn, that these distributions are not determined by the model.

3. Functional models for statistical data

A further important consequence is that different functional models may be employed for the interpretation of any given set of data. As an illustration I use an example previously discussed by N. Cartwright (1979). The example relates to a statistical collection of humans, Ω, and there are three statistical variables: X records whether a person regularly smokes ($X = 1$) or not ($X = 0$); Z records whether a person regularly exercises ($Z = 1$) or not ($Z = 0$); and Y records whether a heart attack occurred ($Y = 1$) or not ($Y = 0$). It will be assumed that the following data (different from Cartwright's) have been observed.

X	Z	$Y = 0$	$Y = 1$	Total
0	0	420	80	500
0	1	95	5	100
1	0	70	30	100
1	1	450	50	500

(6.1)

A simple model construction starts with modal variables $\ddot{X}$, $\ddot{Z}$, and $\dot{Y}$, defined in correspondence to X, Z, and Y, and assumes that the distribution of $\dot{Y}$ depends on the values of $\ddot{X}$ and $\ddot{Z}$:

$$\begin{tikzcd} \ddot{X} \arrow[dr] & \\ & \dot{Y} \\ \ddot{Z} \arrow[ur] & \end{tikzcd} \qquad (6.2)$$

The stochastic function can be written as

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

and, using the conditional frequencies which can be derived from the data in (6.1) as estimates of the conditional probabilities assumed by the model, one gets:[3]

x	z	$\Pr(\dot{Y}=0\,\|\,\ddot{X}=x,\ddot{Z}=z)$	$\Pr(\dot{Y}=1\,\|\,\ddot{X}=x,\ddot{Z}=z)$
0	0	0.84	0.16
0	1	0.95	0.05
1	0	0.70	0.30
1	1	0.90	0.10

(6.3)

Referring to the model, one can now consider a change $\Delta(0,1)$ in the variable $\ddot{X}$ as a functional cause of a stochastic effect. Considering mean differences

$$\Delta^s(\dot{Y};0,1;z) := \mathrm{E}(\dot{Y}\,|\,\ddot{X}=1,\ddot{Z}=z) - \mathrm{E}(\dot{Y}\,|\,\ddot{X}=0,\ddot{Z}=z)$$

as an effect measure, one compares the probabilities of heart attacks of smokers and non-smokers in the covariate context $\ddot{Z}=z$. From the values in (6.3) one finds the following context-depending effects:

$$\Delta^s(\dot{Y};0,1;0) = 0.30 - 0.16 = 0.14$$

$$\Delta^s(\dot{Y};0,1;1) = 0.10 - 0.05 = 0.05$$

Note that the calculation presupposes that a change $\Delta(0,1)$ in $\ddot{X}$ can take place without a simultaneous change of the covariate context. This is an assumption of the model which can not be corroborated from the data.

4. Indeterminate mean effects

This assumption is completely independent of the distribution of the statistical variables corresponding to the exogenous model variables (X and Z in the example). As discussed in § 2, arbitrary distributions of these variables are compatible with the same model. For example, instead of (6.1) the following data could be used:

X	Z	$Y=0$	$Y=1$	Total
0	0	420	80	500
0	1	475	25	500
1	0	350	150	500
1	1	450	50	500

(6.4)

One would get the same conditional probabilities as those shown in (6.3) and consequently also the same context-dependent effects of a change $\Delta(0,1)$ in the variable $\ddot{X}$.

However, since mean effects depend on the distribution of covariate contexts in the respective statistical collections, conclusions of causal effects differ. Using the data (6.1), one is led to

$$P(Y=1 \mid X=1) - P(Y=1 \mid X=0) = \frac{80}{600} - \frac{85}{600} = -0.008$$

Using instead the data in (6.4) one finds:

$$P(Y=1 \mid X=1) - P(Y=1 \mid X=0) = \frac{200}{1000} - \frac{105}{1000} = 0.095$$

In the first case the frequency of heart attacks is lower for smokers than for non-smokers, in the second case it is the other way around.

It should be stressed that these calculations of mean effects relate to the data, not to the functional model. The model (6.2) does not allow one to calculate a mean effect of a change $\Delta(x', x'')$ in the variable $\ddot{X}$ because the effect is context-dependent and different contexts are possible.[4]

5. *Individuals and populations*

There is no unique relationship between functional models and statistical data, if only because functional models are models for generic individuals which refer in an abstract sense to an individual (person, object, or situation). Of course, when there is a population of individuals the model can be applied to each of its members. Still, models for generic individuals can only be used for questions concerning each individual independently. They cannot be used for questions concerning the development of statistical distributions, and relationships between such distributions, in a population. This would require a quite different type of functional model; one would need modal variables which can record statistical distributions and then assume functional relationships between such variables. Models of this kind, here called *population-level models*, will be treated in Chapter 8.

6.2 Experimental and observational data

The distinction between passive observation and active experimentation has a long tradition. This section takes up the distinction in order to discuss some suggestions to connect functional models with data.

1. *Randomized experiments*

Randomized experiments are often assumed to be especially useful to establish causal knowledge. In order to understand the argument, and its limitations, consider a functional model

$$\ddot{X} \longrightarrow\!\!\!\!\rightarrow \dot{Y} \tag{6.5}$$

Suppose that the conditional distributions $\Pr[\dot{Y}|\ddot{X}=x]$ are to be estimated using experimental data. Suppose further that such data can be generated for a collection of individuals (persons, objects, or situations). I represent the collection by $\Omega = \{\omega_1, \ldots, \omega_n\}$. The experimenter can, for each individual $\omega \in \Omega$, decide about a treatment, i.e. a value of $\ddot{X}$, and then observe a value of $\dot{Y}$. At the end there are values of a two-dimensional statistical variable

$$(X, Y) : \Omega \longrightarrow \tilde{\mathcal{X}} \times \tilde{\mathcal{Y}} \tag{6.6}$$

which can be used to calculate (a model function for) the regression function $x \longrightarrow \mathrm{P}[Y|X=x]$. This function can then be considered as an estimate of the function $x \longrightarrow \Pr[\dot{Y}|\ddot{X}=x]$ assumed by the model (6.5).

The data generation is called a *randomized experiment* if the experimenter uses a random generator with a known distribution to decide about the assignment of treatments (values of $\ddot{X}$) to the members of Ω.[5] Why might this be a good idea? One can imagine that $\dot{Y}$ not only depends on values of $\ddot{X}$ but also on values of another variable $\ddot{Z}$. For the sake of argument it will also be assumed that values of $\ddot{Z}$ are fixed before any intervention by the experimenter. The theoretical supposition then refers to a model of the form

$$\ddot{X} \longrightarrow\!\!\!\!\rightarrow \dot{Y} \twoheadleftarrow\!\!\!\!\longleftarrow \ddot{Z} \tag{6.7}$$

Would $\ddot{Z}$ be observable one could use, instead of (6.6), values of a statistical variable $(X, Y, Z) : \Omega \longrightarrow \tilde{\mathcal{X}} \times \tilde{\mathcal{Y}} \times \tilde{\mathcal{Z}}$ and then investigate whether, and how, effects of changes $\Delta(x', x'')$ also depend on values of $\ddot{Z}$. If values of $\ddot{Z}$ cannot be observed one has to consider a model of the form

$$\ddot{X} \longrightarrow\!\!\!\!\rightarrow \dot{Y} \twoheadleftarrow\!\!\!\!\longleftarrow \dot{Z} \tag{6.8}$$

such that (6.5) can be viewed as a reduced version (see Section 4.3). Then two questions arise: Is the assumption made in the formulation of (6.8), that $\dot{Z}$ is stochastically independent of $\ddot{X}$, valid? Which distribution can be assumed for $\dot{Z}$?

Randomization is related to the first question and makes it plausible to use a model of the form (6.8).[6] Completely independent of the randomization one can approach the second question when referring to a population Ω. One can simply stipulate that the distribution of $\dot{Z}$ approximately equals the statistical distribution of Z in Ω: $\Pr[\dot{Z}] \approx \mathrm{P}[Z]$. While it is true (by assumption) that Z cannot be observed, this stipulation nevertheless allows one to explicate the relationship between the original model (6.5) and the enlarged model (6.8) which is used for an interpretation of the available data. Effects of a change $\Delta(x', x'')$ calculated with the reduced model are then understandable as mean effects with respect to the distribution of covariate contexts in the experimental population.

2. What can be achieved with randomization?

The conclusion so far is that randomization can be viewed as a method for the construction of mean effects with respect to the distributions of unobserved covariates. It is important, however, to understand that the mean effects nevertheless depend on the contingent realization of these unobserved distributions in the experimental population that was used to generate the data.

A variant of the school example discussed in § 5 of Section 5.3 will be used to illustrate the argument. $\ddot{X}$ records the school type (1 or 2), and $\dot{Y}$ records the child's educational success (0 or 1); in addition there is a variable $\ddot{Z}$ for the parent's educational level (0 = low or 1 = high).[7] The theoretical model has the form (6.7) and, as in Section 5.3, the following function will be assumed:

z	x	$\mathrm{E}(\dot{Y}\mid\ddot{X}=x,\ddot{Z}=z)$
0	1	0.5
0	2	0.7
1	1	0.8
1	2	0.9

(6.9)

As a precondition for the practicability of a randomized experiment it will be assumed that the children can be distributed among the school types independently of the educational level of their parents. If p_x and $1-p_x$ denote the probabilities for the assignment of school types 2 and 1, respectively, the randomization process results in a statistical variable $X:\Omega\longrightarrow\{1,2\}$ with distribution $\mathrm{P}(X=2)\approx p_x$. If eventually the educational success has been observed, the data then provide, for each child $\omega\in\Omega$, the school type $X(\omega)$ and the educational success $Y(\omega)$. A further statistical variable, Z, records the educational levels of the children's parents, and $p_z := \mathrm{P}(Z=1)$ denotes the proportion of children in Ω whose parents have a high educational level.

The task is to compare, and interpret, the mean values $\mathrm{M}(Y|X=2)$ and $\mathrm{M}(Y|X=1)$ which can be calculated from the data. As a result of the randomization, one can assume that X and Z are approximately independent (in a statistical sense) and, referring to model (6.8), this allows one to use the following approximation:

$$\mathrm{M}(Y|X=x)\approx \mathrm{E}(\dot{Y}|\ddot{X}=x,\dot{Z}=1)p_z+\mathrm{E}(\dot{Y}|\ddot{X}=x,\dot{Z}=0)(1-p_z)$$

Using then the values assumed in (6.9), one finds

$$\mathrm{M}(Y|X=2)-\mathrm{M}(Y|X=1)\approx 0.2-0.1p_z$$

Obviously, the effect of the school type for a child's educational success also depends on the distribution of the parent's educational levels as realized in the population of children selected for the experiment.

3. The standard argument for randomization

An often evoked argument for randomization is as follows:

> 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 and Winship 2002: 433)

At first sight the idea seems plausible. In order to discover the effect, defined for some variable $\dot{Y}$, of a change $\Delta(x_1, x_2)$ in a variable $\ddot{X}$ one needs data for correspondingly defined statistical variables X and Y (which relate to some population Ω) that allow the comparison of two groups:

$$\Omega_2 := \{\omega \,|\, X(\omega) = x_2\} \quad \text{and} \quad \Omega_1 := \{\omega \,|\, X(\omega) = x_1\}$$

Referring to a (randomized) experiment, Ω_2 is the treatment and Ω_1 the control group. It seems plausible: If values of X result from a randomization device, the two groups can be assumed to be similar with respect to all variables having fixed values before the randomized assignments took place. However, why should this be important for the interpretation of the finally found effect of $\Delta(x_1, x_2)$?

Think of a variable Z with possibly different distributions in Ω_1 and Ω_2. There are two possibilities. Either Z is irrelevant for the effect; then a randomization with respect to this variable is not necessary. Or the effect depends on values of Z. If the experimental population is sufficiently large one may assume that, due to the randomization, X and Z are approximately independent (in a statistical sense): $\mathrm{P}[Z|X = x_1] \approx \mathrm{P}[Z|X = x_2]$. This allows the expression of the effect of $\Delta(x_1, x_2)$ as a simple difference of means:

$$\mathrm{M}(Y|X = x_2) - \mathrm{M}(Y|X = x_1) \approx \Delta^a(Y|x_1, x_2; Z) \tag{6.10}$$

where the right-hand side is defined by

$$\Delta^a(Y|x_1, x_2; Z) := \sum_z (\mathrm{M}(Y|X = x_2, Z = z) - \mathrm{M}(Y|X = x_1, Z = z))\,\mathrm{P}(Z = z) \tag{6.11}$$

However, (6.10) does not imply that the mean effect calculated from the randomized data is independent of Z and only attributable to the change $\Delta(x_1, x_2)$. As shown by the example in § 2, the effect can also depend on the distribution of Z in the experimental population that generated the data; and the effects in a population with another distribution of Z might be quite different. In fact, supposing that the effect also depends on values of Z, reliable conclusions can only be based on a knowledge of the context-specific effects

$$\mathrm{M}(Y|X = x_2, Z = z) - \mathrm{M}(Y|X = x_1, Z = z)$$

and this knowledge, of course, can only be gained from additional data for the variable Z.[8]

4. Limitations of using mean effects

The argument can be continued: If it would be possible to observe values of Z it would not be necessary to randomize with respect to this variable. It would be possible, then, to investigate how the effect also depends on values of Z. Of course, without a randomization there is no guarantee that X and Z will be approximately independent. However, why should this be relevant for the definition of a causal effect?

The question concerns a problem that, in a sense, results from a theoretical conflict. On the one hand, one wants to estimate a definite effect but, on the other hand, one knows, or believes, that the effect is context-dependent. In the example, the effect of $\Delta(x_1, x_2)$ also depends on values of Z. In order to attribute a definite effect to $\Delta(x_1, x_2)$ one has either to refer to a specified covariate context $Z = z$, or one has to rely on a mean effect defined with respect to some distribution of covariate contexts. Viewed in this way, (6.11) is a proposal for the definition of a mean effect.

However, mean effects can be determined in many different ways. Why use (6.11)? A possible consideration refers to an individual ω randomly drawn from Ω. If one does not know the covariate context $Z(\omega)$, but only knows that ω is a member of Ω, and Z is distributed in Ω according to $\mathrm{P}[Z]$, it might be reasonable to calculate an expectation as in (6.11). However, one can easily imagine situations where the available information suggests otherwise. For example, assume that 40 per cent of the children in the population Ω have parents with a higher educational level: $\mathrm{P}(Z = 1) = 0.4$, and also assume that the assignment to school types depends in the following way on the parents' educational level:

$$\mathrm{P}(X = 2 \,|\, Z = 0) = 0.4 \quad \text{and} \quad \mathrm{P}(X = 2 \,|\, Z = 1) = 0.8$$

Using the data from (6.9), one then finds for the mean effect defined in (6.11) the value $(0.7 - 0.5)0.6 + (0.9 - 0.8)0.4 = 0.16$. Now, referring to a child randomly drawn from Ω, one might expect the probability of an educational success to be about 16 percentage points higher in school type 2 than in school type 1.

However, there are other and possibly more interesting questions. For example, how would the probability of an educational success change if a child who began in school type 1 changes into school type 2? To answer this question one would need to randomly draw the child from the subpopulation $\Omega_1 := \{\omega | X(\omega) = 1\}$ in which the distribution of the parents' educational levels is given by

$$\mathrm{P}(Z = 0 \,|\, X = 1) = 0.82 \quad \text{and} \quad \mathrm{P}(Z = 1 \,|\, X = 1) = 0.18$$

and then one finds a different mean effect, namely

$$(0.7 - 0.5)0.82 + (0.9 - 0.8)0.18 = 0.18$$

5. Arguments based on potential outcomes

A somewhat different version of the standard argument for randomization has been proposed by authors who adhere to the potential outcomes approach (see Section 5.2). In order to discuss this version of the argument, I again follow Holland's exposition. Using the notations of Section 5.2, the following table can serve to explain the argument:

$$\begin{array}{ccccc} \omega & X(\omega) & Y_0(\omega) & Y_1(\omega) & Y(\omega) \\ \hline \omega_1 & x_1 & y_{01} & y_{11} & y_1 \\ \omega_2 & x_2 & y_{02} & y_{12} & y_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \omega_n & x_n & y_{0n} & y_{1n} & y_n \end{array} \tag{6.12}$$

The table relates to individual units in a population Ω. To each individual ω_i can be applied one of the treatments $X = 0$ or $X = 1$. $Y_x(\omega_i)$ is the outcome that would result if the treatment $X = x$ had been applied to ω_i.[9] $X(\omega_i)$ and $Y(\omega_i)$ record the treatment actually applied and the actual outcome, respectively:

$$Y(\omega_i) = \begin{cases} Y_0(\omega_i) & \text{if } X(\omega_i) = 0 \\ Y_1(\omega_i) & \text{if } X(\omega_i) = 1 \end{cases}$$

The goal of the experiment is to calculate a mean effect

$$\mathrm{M}(Y_1 - Y_0) = \textstyle\sum_{i=1,n} (Y_1(\omega_i) - Y_0(\omega_i))/n$$

Since only one treatment can be applied to each individual, how can this mean effect be discovered? The argument as given, for example, by Holland (1986: 948–9) relies on a randomization of the assignment of treatments and can be summarized as follows: If values of X are randomly assigned, then

$$\Omega_0 := \{\omega \,|\, X(\omega) = 0\} \quad \text{and} \quad \Omega_1 := \{\omega \,|\, X(\omega) = 1\}$$

are simple random samples from Ω and, given the samples are sufficiently large, one gets estimates of the potential outcomes as $\mathrm{M}(Y_0) \approx \mathrm{M}(Y|X = 0)$ and $\mathrm{M}(Y_1) \approx \mathrm{M}(Y|X = 1)$, and consequently also a reasonable estimate of the mean difference between the potential outcomes:

$$\mathrm{M}(Y_1 - Y_0) \approx \mathrm{M}(Y|X = 1) - \mathrm{M}(Y|X = 0)$$

The argument obviously depends in a crucial way on the assumption that values of the variables Y_0 and Y_1 exist in some sense already before any specific treatment is applied. It is this assumption that allows the thinking of randomization as if it were a method of generating a random sample of both.

6. Causal and pseudo-descriptive questions

The argument of the potential outcome approach can be illustrated using an urn model.[10] For each individual ω_i, the urn contains a ball one half blue-colored and the other half red-colored; on the blue-colored half y_{0i} is written, on the red-colored half it is y_{1i}. Then two random samples are drawn; the first sample provides values of the blue halves and the second sample provides values of the red halves of their balls, respectively.

This view is, however, in conflict with the meaning of a causal question which presupposes the existence of some process that leads from a cause to an effect. While more detailed conceptualizations of such processes also depend on whether one considers dynamic or comparative causes, the essential idea can be explained if one simply assumes values of a variable X being conditions of values of another variable Y:

$$X(\omega_i) \xrightarrow{\text{causal process}} Y(\omega_i)$$

The crucial point is that the potential outcome approach does not refer to a causal process but presupposes a selection process instead. Again, the school example can serve for an illustration. $X(\omega_i)$ determines the school type which then is a condition for a subsequent process that brings about a value of $Y(\omega_i)$, the child's educational success. Obviously, it is important that one can think of a process for which $X(\omega_i)$ is a relevant condition. This is of particular importance if there are further conditions which might become effective after the realization of the condition $X(\omega_i)$. To continue with the example, it will be assumed that the child's educational success also depends on whether his or her parents provide much support ($X' = 1$) or otherwise ($X' = 0$). Following the potential outcome approach one would need to define further fictitious variables $Y'_{x'}$ which capture the child's educational success if $X' = x'$. As an extension of (6.12) one would get the following extended scheme:

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}

This extended scheme obviously excludes the possibility to vary X and X' independently. Referring to a child ω_i, if the school type x_i is fixed, the educational success $Y_{x_i}(\omega_i)$ is already assumed to exist, whatever the value of $X'(\omega_i)$ might be.[11] Imagine, for example, the following values:

$$y_{0i} = 0, \quad y_{1i} = 1, \quad y'_{0i} = 0, \quad y'_{1i} = 1$$

If $x_i = 1$ and $x'_i = 0$, should it be concluded that the educational success simultaneously is 1 and 0? A better conclusion is that the potential outcome approach is, in particular, not suited for the consideration of causal processes which depend on several causes.[12]

The potential outcome approach replaces a causal question by a pseudo-descriptive question. This might be reasonable if one refers to data, i.e. to causes and effects already realized for the members of some population Ω. Then one can consider a counterfactual question: Which values of Y might have been realized if, instead of the actually realized values $X(\omega)$, some other values would have been realized? However, even if it were possible to find, and justify, an answer, the answer would only be relevant for Ω without any implications for individuals, or situations, not belonging to Ω.

7. Observational data

So far only data were considered that result from experiments in which an experimenter has some control about (the assignment of) some of the conditions on which values of an outcome variable depend. When no control can be exercised and values of the relevant variables are merely observed, one commonly speaks of *observational data*. This is the kind of data almost always used in statistical social research. Since the data do not arise from (randomized) manipulation of conditions, many authors judge such data as problematic for causal questions.[13] For example, Rosenbaum (1984: 41) writes:

> 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.

Of course, difficulties only arise if observational data are to be used for causal questions and not just for descriptive purposes. It is important, however, that causal questions in the realm of social research can be conceptualized quite differently. One might ask how observational data can be used to estimate causal effects defined in the framework of randomized experiments. One then views observational data from the point of view of an only hypothetically possible (randomized) experiment.[14] However, this view is of only limited use in social research.

To elucidate the argument, assume that one is interested in a functional model $\ddot{X} \twoheadrightarrow \dot{Y}$ which refers to a hypothesized relationship between school type ($\ddot{X}$) and educational success ($\dot{Y}$). The model formulation leaves indeterminate how values of $\ddot{X}$ come into being; one may think of a theoretical context as follows:

$$\begin{array}{ccc} \boxed{\text{substantial process}} & \longrightarrow & \boxed{\text{values of } \ddot{X}} \\ & & \vdots \\ & & \boxed{\text{observations}} \end{array} \tag{6.13}$$

There is a substantial process that generates the empirical facts to be captured by values of $\ddot{X}$, and afterwards an observational process possibly generates data (about previously existing facts). In the example, the school type is determined by a substantial process for each child. The data generating process provides knowledge about the realized attachments. The argument shows that randomization is not a data generating, but a substantial process.

How can one attribute a determinate effect to a cause $\Delta(x', x'')$ in the variable $\ddot{X}$? At first sight it seems that it is irrelevant how the cause $\Delta(x', x'')$ comes into being.[15] However, this is only true if the process that generated the cause is not already in some relevant sense context-dependent on conditions which also have implications for the supposed effect. Suppose that the child's educational success also depends on the parent's educational level, $\ddot{Z}$. One can then consider the following possibilities:

$$
\text{(a)}\quad
\begin{tikzcd}
\ddot{Z} \arrow[dr, two heads] & \\
& \dot{Y} \\
\ddot{X} \arrow[ur, two heads] &
\end{tikzcd}
\qquad\qquad
\text{(b)}\quad
\begin{tikzcd}
\ddot{Z} \arrow[dr, two heads] \arrow[dd, two heads] & \\
& \dot{Y} \\
\dot{X} \arrow[ur, two heads] &
\end{tikzcd}
\qquad (6.14)
$$

While model (a) does not make an assumption about the process that creates values of $\ddot{X}$, model (b) assumes that this process also depends on the parents' educational level (implying the substitution of $\ddot{X}$ by a stochastic variable $\dot{X}$). Model (b) is therefore not compatible with the idea of a randomization with respect to $\dot{X}$ and, consequently, one cannot use this idea in order to define the effect of a change in $\dot{X}$.

It is important to see that the problem does not result from the impossibility of a randomization but from its being theoretically incompatible with the processes as they may develop in the empirical world.[16] Even if it would be possible in some experiment to assign children randomly to different school types, the result would not be relevant because the process to be clarified by the causal analysis is different from the one generated by the randomized experiment.

In conclusion, definitions of causal effects based on randomized experiments cannot, without further qualifications, be viewed as a theoretical ideal. To the contrary, only an extensive elaboration on contexts and purposes will make knowledge gained from randomized experiments useful. This conclusion is of particular importance if the causal knowledge is intended to serve as an assessment of possible courses of action; since then it must be taken into account which actors actually perform the decisions (e.g. decide about school types of children).

6.3 Interventions and reference problems

The previous section argued that the idea of randomized experiments not only is often impractical but runs into theoretical problems when applied to social processes. Additional problems arise from the involvement of human actors.

1. Causal knowledge and interventions

It is obvious that causal knowledge often serves to assess possible effects of *interventions* (a term here used to denote deliberately generated causes). Some authors have tried to make this a leading idea for a definition of causal relationships.[17] The idea seems particularly attractive in the context of experiments where causes can be generated deliberately. However, in considerations of social processes one has to take into account that they already involve actors, and that these actors perform interventions and thereby generate effects.

Before elaborating on this point, it is worth remembering that the idea of a (fictitious) intervention can be applied without difficulties only to the exogenous variables of a model. The very definition of exogenous variables requires that they may be given arbitrary values. In contrast, the values of endogenous variables depend on values of other variables. The models in (6.14) may serve to illustrate the difference. Model (a) represents the school type by an exogenous variable $\ddot{X}$ so that arbitrary assignments are possible. In contrast, in model (b) $\dot{X}$ is an endogenous variable whose values depend on another variable, $\ddot{Z}$, the parents' educational level. What is implied by assignments of values to $\dot{X}$ is therefore not immediately clear.

Some authors have proposed that by stipulating an intervention with respect to some variable, that variable is made automatically into an exogenous variable;[18] for example 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.

The idea is to use a modified model in order to make a *hypothetical experiment* (Woodward 1999: 201) conceivable. In order to investigate possible effects of a change $\Delta(x', x'')$ in an endogenous variable $\dot{X}$, a new model is constructed that no longer contains the function that made $\dot{X}$ dependent on other variables in the original model.[19] For example, starting from model (6.14b) a hypothetical experiment intervening in $\dot{X}$ would presuppose model (6.14a) which contains the exogenous variable $\ddot{X}$ instead of the endogenous variable $\dot{X}$.[20]

2. Causes as parts of processes

The proposal just described evades an answer to the question of a possible causal meaning of endogenous variables. In fact, the two model versions in (6.14) correspond to different questions. Model (a) relates to the question how a child's educational success depends on the parents' educational level and on the school type, and deliberately leaves unspecified how values of these two "independent" variables come into being. Model (b), on the other hand, explicitly assumes a process by which the choice of a school type depends on the parents' educational level. While in both models the child's educational success depends on the parents' educational level *and* on the school type, one needs to add that the school type has no longer an autonomous causal meaning in model (b).

The change of status of variables in models (a) and (b) corresponds to a change of questions that may be answered by the use of the models. More often than not, interest centers on interventions which actually can be performed. One can speak of *praxeological questions* that ask for interventions by which people can achieve specified effects.[21] The question how a child's educational success comes into being is of another kind. This question asks for a causal explanation; the goal is some insight into a process that generated a specific fact (singular explanation) or into processes which generate facts of some kind (generic explanation). In any case it becomes necessary to place supposed causes into the context of a substantial process that leads to the possible effects. This also might require to refer to actors; but their interventions are then to be seen as being part of the substantial process to which the model relates.

3. Different kinds of interventions

The reference to interventions in connection with models is fictitious: The creator, or user, of the model assigns values to some variables defined in the model. Whether, and how, one can speak of real interventions primarily depends on how the model relates to the empirical world. We are here concerned with models of social processes that necessarily involve human actors. Two kinds of interventions must therefore be distinguished:[22]

- *Substantial actors* involved in the substantial processes that generate the facts referred to by the model's variables. In the school example these are first of all the children and their parents who, *inter alia*, decide on the school type; further actors (e.g. teachers) are then involved in the process that subsequently leads to the pupils' success or failure.
- *Secondary actors* which are (implicitly) referred to if one speaks of observations, the construction of models, and of real and fictitious experiments. One can think, e.g. of researchers, administrators, and advisory boards.

Note that these need not be different groups of actors. The distinction rather results from the fact that "one" can, empirically as well as theoretically, *refer to* processes which involve actors. "One" then, by performing this kind of reference, becomes a secondary actor, and the actors referred to become the substantial actors.

This distinction between kinds of actors leads to a corresponding distinction between kinds of interventions. On the one hand are the interventions performed by substantial actors; these will be called *substantial interventions*. On the other hand are the inventions performed by secondary actors; these will be called *modal interventions* (because they use models for the reasoning about modal questions). This distinction suggests a reasonable interpretation of modal interventions. If model variables are exogenous one may make assumptions about their values.

However, this is impossible without contradicting the model when variables are endogenous, because then the model already implies the existence of functions that generate values of the variables. Instead, modal interventions must then be applied to these functions.

4. Assumptions about people's behavior

Possible interpretations of modal interventions concerning the functions of a model depend on the processes to which the functions relate. Functions in models for social processes often relate to the behavior (or implications of the behavior) of the substantial actors involved in the processes. Modal interventions then essentially equal assumptions about the behavior of these actors.

Again, the school example may serve as an illustration. In this example, the relevant behavior function connects the parents' educational level with the choice of a school type:

$$z \longrightarrow \Pr[\dot{X} \mid \ddot{Z} = z] \tag{6.15}$$

One question concerns the effect on children's educational success of a change, or difference, in the values of $\ddot{Z}$, the parents' educational level, being mediated through the school type. Using the abbreviations

$$p_{x,0} = \Pr(\dot{X} = 2 \mid \ddot{Z} = 0) \quad \text{and} \quad p_{x,1} = \Pr(\dot{X} = 2 \mid \ddot{Z} = 1)$$

one gets from the data in (6.9):

$$\mathrm{E}(\dot{Y}|\ddot{Z} = 1) - \mathrm{E}(\dot{Y}|\ddot{Z} = 0) = 0.3 + 0.1\,p_{x,1} - 0.2\,p_{x,0} \tag{6.16}$$

The difference between the probabilities of an educational success becomes greater if more children of parents with a higher educational level visit school type 2, and the difference becomes smaller if more children of parents with a lower educational level visit school type 2.

A complementary question concerns the effect on children's educational success of a change, or difference, in the school type, taking now into account that the composition, with respect to parents' educational level, is different in the school types. In addition to assumptions about the behavior function (6.15), one therefore needs assumptions about the distributions of the parents' educational levels. Using in model (6.14b) instead of $\ddot{Z}$ a stochastic variable $\dot{Z}$, one can, with the help of the formula

$$\mathrm{E}(\dot{Y}|\dot{X} = x) = \frac{\sum_z \mathrm{E}(\dot{Y} \mid \dot{X} = x, \dot{Z} = z)\Pr(\dot{X} = x \mid \dot{Z} = z)\Pr(\dot{Z} = z)}{\sum_z \Pr(\dot{X} = x \mid \dot{Z} = z)\Pr(\dot{Z} = z)}$$

calculate an effect

$$\mathrm{E}(\dot{Y} \mid \dot{X} = 1) - \mathrm{E}(\dot{Y} \mid \dot{X} = 0) \tag{6.17}$$

Assuming once again $p_z = \mathrm{P}(Z = 1) = 0.4$, Figure 6.1 demonstrates the dependence of the mean effect (6.17) on the probabilities $p_{x,0}$ and $p_{x,1}$. The figure

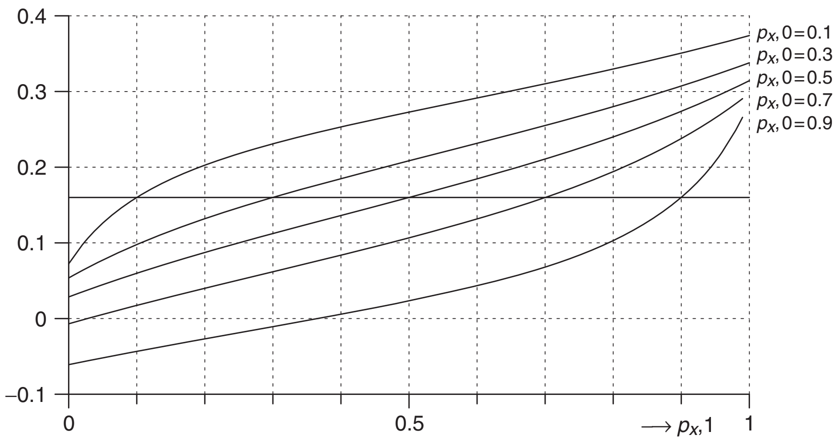


Figure 6.1 For $p_{x,0} = 0.1, 0.3, 0.5, 0.7$, and 0.9, the curves show how the effect $\mathrm{E}(\dot{Y}|\dot{X} = 1) - \mathrm{E}(\dot{Y}|\dot{X} = 0)$ depends on the probabilities $p_{x,1}$ (shown on the X-axis).

also shows the mean effect 0.16 that would result from a randomized distribution ($p_{x,0} = p_{x,1}$).[23]

Notice that (6.16), as well as Figure 6.1, refers to a combined effect that results from different educational levels of the parents, from different school types, and from the children's distribution on the school types. However, exactly such interplay of conditions must be investigated in order to understand how differences in children's educational success come into being.

5. *Relevance of the substantial actors*

Authors who conceive of causality in terms of interventions often distinguish between active and passive predictions; for example:

> 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.
>
> (Woodward 2003: 47)

A standard example is using a barometer. Observing a change in the atmospheric pressure this information can be used to predict some change of the weather; on the other hand, intentionally changing the reading of a barometer will not change the weather. However, Woodward's argument needs modification if variables obtain their values by actions of substantial actors. Passive observations, then, nevertheless provide information about results of interventions, and it becomes important to distinguish these substantial interventions from the modal interventions of the model builder.

Also, when processes involve substantial actors, the randomization idea becomes problematic because randomization then substitutes realizations of a random generator for the decisions of the substantial actors. In consequence, the belief that randomization is a desirable method[24] either presupposes a process not involving actors capable of acting on their own decisions, or it deliberately ignores the actors' capabilities.

A further implication is of some importance: It is often reasonable to assume that the substantial actors in social processes have better knowledge about relevant context conditions than the secondary actors who try to get some insight into these processes with the help of models.[25] One may then assume that there is at least one variable that is unobserved by the secondary actors but accessible to the substantial actors. Some authors suggest that in these conditions estimates of causal effects are always "biased."[26] The suggestion is questionable, however, because there is no model that could establish an "unbiased" definition. In fact, almost always functional models only allow one to explicate generic causal effects having realizations depending on context conditions not explicitly represented in the model.

6. *Distortions due to self selection?*

The belief that self selection can lead to "biased" estimates (of causal effects) is widespread in the econometric literature.[27] For a proper understanding the following distinctions must be kept in mind:

(a) The selection problem can refer to a data generating process as depicted in the following diagram:

$$Y \xrightarrow{\text{data generating process}} Y^*$$

One is interested in the distribution $\mathrm{P}[Y]$ of a statistical variable Y; but the data generating process provides data only about a variable Y^* having a distribution $\mathrm{P}[Y^*]$ which is different from $\mathrm{P}[Y]$. Using then the data to estimate quantities defined with respect to $\mathrm{P}[Y]$ produces biased estimates.[28] In this context one can speak of self selection if the data generating process depends on decisions of substantial actors. For example, assume that a survey is performed in order to get information about the income distribution of households, $\mathrm{P}[Y]$. Then self selection problems can arise because values of Y^* produced by the survey depend on how the interviewed persons respond on the income question.

(b) When one is interested in causal processes which depend on decisions of substantial actors the problem takes a different form. In the school example self selection refers to the fact that values of the school type variable result from decisions of the children and their parents. This, however, is not a data generating process (which may be viewed as producing distorted data), but a substantial process that produces part of the causes which are effective in the real world.

Also in the school example, self selection (school type decisions of children and their parents) may be viewed as creating distorted samples: Starting with variables Z (parent's educational level), X (school type), and Y (children's educational success), defined for some population Ω, self selection leads to different distributions $\mathrm{P}[Z|X = 0] \neq \mathrm{P}[Z|X = 1]$. Using then, for example, data from the children in only one school type, the resulting distribution of Z would be a biased estimate for $\mathrm{P}[Z]$.

However, the question how to attribute a specific causal effect to a difference $\Delta(x', x'')$ in the school type variable does not translate into an equivalent estimation problem. If one decides to estimate the effect $\Delta^a(Y|x', x''; Z)$ defined in (6.11), using the difference of means, $\mathrm{M}(Y\,|\,X = x'') - \mathrm{M}(Y\,|\,X = x')$, would lead to a biased estimate. However, why estimate $\Delta^a(Y|x', x''; Z)$? In fact, knowledge about this version of a mean effect is not required for the insight that a child's educational success depends on the school type and the parents' educational level. Knowing the mean effect, in particular, provides no hints about the causal roles played by the two conditions. This would require estimating separately the two functions which characterize the causal process in the model (6.14b).

7. Selection problems from expectations?

The econometric literature that deals with questions of self selection often argues with expectations. For an illustration I use an example discussed by Maddala (1977):

> 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 into 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
>
> $D = 1$ if the individual goes to college
>
> $\;\;\;= 0$ otherwise
>
> and estimate an earnings function with D as an extra explanatory variable. This, however, is not a satisfactory solution to the problem.
>
> (p. 351)

In trying to provide reasons for this statement, Maddala refers to a situation without covariates so that (in his notation) $\bar{y}_1 - \bar{y}_2$ is the mean income difference between persons with and without a college education:

> [...] 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 $\bar{y}_1 - \bar{y}_2$.

But how can a selection problem result from the fact that people use expectations for their decisions?[29] To begin with, think of a model

$$\ddot{D} \twoheadrightarrow \dot{Y}$$

that corresponds to the assumption that the income, $\dot{Y}$, depends on the level of education, $\ddot{D}$ ($\ddot{D} = 1$ if the person has a college education, $\ddot{D} = 0$ otherwise). This model cannot, however, lead to a selection problem because (given Maddala's preconditions) all persons would make their decision in the same way, depending only on $\mathrm{E}(\dot{Y}|\ddot{D}=1) - \mathrm{E}(\dot{Y}|\ddot{D}=0)$ being greater or less than zero. However, this value is fixed within the model. The decision for or against a college education must therefore depend on values of at least one further variable, say $\ddot{Z}$, that further differentiates between background conditions. One may consider two versions:

$$\text{(a)}\quad \begin{tikzcd} \ddot{Z} \arrow[d, two heads] & \\ \dot{D} \arrow[r, two heads] & \dot{Y} \end{tikzcd} \qquad\qquad \text{(b)}\quad \begin{tikzcd} \ddot{Z} \arrow[d, two heads] \arrow[dr, two heads] & \\ \dot{D} \arrow[r, two heads] & \dot{Y} \end{tikzcd} \tag{6.18}$$

In version (a), only the decision variable $\dot{D}$ depends on $\ddot{Z}$. Persons with and without a college education may then be different with respect to $\ddot{Z}$, but this would be irrelevant for the relationship between $\dot{D}$ and $\dot{Y}$. Consider then version (b). This version implies that $\ddot{Z}$ and $\dot{D}$ are interactive conditions for $\dot{Y}$; there are at least two values, z' and z'', such that

$$\mathrm{E}(\dot{Y}|\dot{D}=1, \ddot{Z}=z') > \mathrm{E}(\dot{Y}|\dot{D}=0, \ddot{Z}=z')$$
$$\mathrm{E}(\dot{Y}|\dot{D}=1, \ddot{Z}=z'') < \mathrm{E}(\dot{Y}|\dot{D}=0, \ddot{Z}=z'')$$

because otherwise all decisions must be either for or against a college education. However, then there is no uniquely defined causal effect that might serve as the target of an (unbiased) estimation.

8. Fictitious stochastic anticipations

The argumentation in the previous paragraph was based on the idea that an expectation must be connected, in some way, with information that is available at the time the expectation is made. It was therefore assumed that there is a further variable, $\ddot{Z}$, being of some relevance for the subsequent income $\dot{Y}$, and that all substantial actors know their value of $\ddot{Z}$, allowing them to make their expectations, and decisions, dependent on that knowledge. However, in the cited example, Maddala follows a different route and assumes that expectations can directly, without any real or imagined informational basis, refer to incomes that might be achieved with or without a college education.

This then leads to a variant of the potential outcome approach (see Section 6.2, § 5) that uses stochastic individual-level effects:

$$\dot{D} \longrightarrow\!\!\!\!\rightarrow (\dot{Y}_0, \dot{Y}_1) \tag{6.19}$$

The stochastic variables record the income that could be achieved with ($\dot{Y}_1$) and without ($\dot{Y}_0$) a college education. It is assumed, then, that each individual can have expectations about his or her values of these variables and that decisions for or against a college education can be based on these expectations. These assumptions imply, however, that one no longer refers to the model (6.19), but to a different model

$$(\dot{Y}_0, \dot{Y}_1) \longrightarrow \dot{D} \longrightarrow\!\!\!\!\rightarrow \dot{Y} \tag{6.20}$$

that contains a deterministic function, connecting ($\dot{Y}_0$, $\dot{Y}_1$) with $\dot{D}$, specified by the decision rule

$$\dot{D} = \begin{cases} 1 & \text{if } \dot{Y}_1 \geq \dot{Y}_0 \\ 0 & \text{otherwise} \end{cases} \tag{6.21}$$

Obviously, $\dot{Y}_0$ and $\dot{Y}_1$ are now latent variables without any definite (causal) relationship with the income variable $\dot{Y}$. Instead, like a secret oracle, they only provide imaginations for the decision for or against a college education. Maddala therefore adds the assumption that, in the mean, expectations become true:[30]

$$\begin{aligned} \mathrm{E}(\dot{Y}|\dot{D}=1) &= \mathrm{E}(\dot{Y}_1|\dot{D}=1) \\ \mathrm{E}(\dot{Y}|\dot{D}=0) &= \mathrm{E}(\dot{Y}_0|\dot{D}=0) \end{aligned} \tag{6.22}$$

So he finally arrives at a complete model (In the graphical representation in (6.20) one may add a double arrow that directly connects ($\dot{Y}_0$, $\dot{Y}_1$) and $\dot{Y}$) demonstrating a purported selection problem:

$$\mathrm{E}(\dot{Y}|\dot{D}=1) - \mathrm{E}(\dot{Y}|\dot{D}=0) \neq \mathrm{E}(\dot{Y}_1) - \mathrm{E}(\dot{Y}_0) \tag{6.23}$$

It suffices to assume that the distributions of $\dot{Y}_0$ and $\dot{Y}_1$ to some extent overlap (otherwise always $\dot{D}=0$ or $\dot{D}=1$).[31]

An obvious objection is that the model does not allow for a causal interpretation: The assumed stochastic function $\dot{D} \longrightarrow\!\!\!\!\rightarrow \dot{Y}$ does not refer to a causal process.

This corresponds to the critique of the potential outcome approach discussed in Section 6.2 (§ 6). Maddala's model can also be used to illustrate the arbitrariness of arguing with distortions when referring to (supposed) causal effects. It suffices to assume that, according to the model, there is a realizable process (a process which can be simulated):

(1) In a first step, random generators are used to create values y_0 and y_1 of the stochastic variables $\dot{Y}_0$ and $\dot{Y}_1$, respectively.
(2) Then, if $y_1 \geq y_0$, one decides for a college education and receives the income $\dot{Y} = y_1$; otherwise one decides against a college education and receives $\dot{Y} = y_0$.

This allows for the definition of the effect of a difference $\Delta(0, 1)$ in the variable $\dot{D}$ by

$$\Delta^s(\dot{Y}; 0, 1) := \mathrm{E}(\dot{Y}|\dot{D} = 1) - \mathrm{E}(\dot{Y}|\dot{D} = 0) \tag{6.24}$$

which equals the left-hand side of (6.23). Why might it be important that this definition differs from the other one?:

$$\Delta^e(\dot{Y}; 0, 1) := \mathrm{E}(\dot{Y}_1) - \mathrm{E}(\dot{Y}_0) \tag{6.25}$$

This definition actually corresponds to another model that assumes an exogenous variable $\ddot{D}$ instead of the endogenous variable $\dot{D}$. This would allow one to perform an intervention as described by Woodward or Pearl.[32] However, if one refers to this modified model, both effect definitions become equivalent; the reasoning therefore does not provide an argument in favor of the definition (6.25).

The important question is, rather, which definition would be adequate for the model (6.20) that contains $\dot{D}$ as an endogenous variable? Since values of $\dot{D}$ result from decisions of substantial actors, one has to think of modal interventions as discussed in § 4. In the current example, this necessitates thinking about causal implications of different decision rules assumed to be followed by the substantial actors. However, this can only be done by using the effect definition (6.24); the alternative definition (6.25) already implies one specific decision rule, namely randomization, and consequently contradicts the assumption of any other decision rule for the substantial actors.

9. *Presuppositions about invariances*

Authors who think of causal effects as resulting from (hypothetical) interventions often propose a further requirement: that, in order to allow for a causal interpretation, the functions of a model should be invariant with respect to possible interventions.[33] This requirement leads, however, into two difficulties. The first one is simply due to the fact that the invariance requirement relates to applications of a model and therefore cannot be formulated without referring to a specified application context.

More important is a second difficulty that results if models relate to substantial processes involving actors. In most cases, then, at least one model function depends on the behavior of substantial actors and those functions obviously cannot be invariant with respect to substantial interventions. However, it also becomes questionable, then, to relate the invariance requirement to modal interventions (of secondary actors) since these interventions consist in modifying assumptions about functions which refer to the behavior of substantial actors.

7 Models with event variables

7.1 Situations and events
1. Events and event types
2. Event variables
3. Referring to event data
4. Functional event models
5. Models for single situations
6. Consecutive situations
7. Ramification of situations

7.2 Event models with time axes
1. Time axes for situations
2. Time-dependent event probabilities
3. Non-occurring events
4. Aggregation of temporal locations
5. Situations with several event variables
6. Static and dynamic contextual variables
7. Time axes for consecutive situations

7.3 Dynamic causality
1. Introductory remarks
2. A definition of dynamic causality
3. Illustration with a random generator
4. Two kinds of modal comparisons
5. Exogenous intervening causes
6. Endogenous intervening causes
7. Concurrent causes
8. Time-dependent dynamic effects
9. Local and integrated effects

Having discussed functional models and notions of functional causality in the three preceding chapters, this chapter continues with the consideration of processes. Several approaches are possible. A first alternative concerns the representation of time. One can use an implicit representation of time by interpreting the functions of a functional model as implying a temporal relationship. Alternatively, one can use a time axis to define processes as sequences of variables (e.g. time series or statistical processes). A second alternative concerns whether processes are defined in terms of states or events. Time series and statistical processes are almost always defined in terms of state variables (see the definitions in Section 3.2). Alternatively, one can think of processes as temporally structured series of events. A formal representation requires event variables which are conceptually different from state variables. A third alternative concerns the distinction between individual-level and population-level processes. Individual-level processes relate to a single object (or situation) considered as a whole. Population-level processes, on the other hand, are conceptual constructions derived from two or more individual-level processes and therefore permit a distinction between two conceptual levels.

Given these alternatives, one can understand the functional models introduced in Chapter 4 as models for individual-level processes defined in terms of state variables (with or without an explicit representation of time). The present chapter, correspondingly, considers functional models for individual-level processes defined in terms of event variables. The first section introduces the conceptual framework and discusses simple models without an explicit time axis. Time axes will be introduced in the second section. This allows one to make event probabilities dependent on temporal durations and on time-varying values of covariates. The third section discusses some possibilities of using event models in order to define a version of dynamic causality that conceives of causes as events and of effects as changes of event probabilities.

7.1 Situations and events

1. Events and event types

A few examples of events suggest an understanding: a child is born, a person becomes employed or unemployed, two persons marry or become divorced, a traffic accident happens. While a general definition is fraught with difficulties, the examples suggest at least the following characteristics: Events are occurrences which can be empirically identified; each event involves at least one object which undergoes a change while the event occurs; events need some time to occur and therefore have a temporal extension; when referring to two events it is often possible to say that one event occurred earlier, or later, than another event; events can be characterized as examples of event types.

Of particular importance is the distinction between events and event types. An event is an empirically identifiable occurrence, for example, the event that two particular persons marry. A corresponding event type would be a possible event whose happening would imply that two persons become married. While an event can only occur once, there can be many events that are examples of the same event type. Characterizing an event by an event type does not, therefore, provide an identifying description. Obviously, such a description would only be possible after the event actually has occurred. On the other hand, whenever one wants to speak of *possible* events (which have not yet occurred, or which might have occurred) the notion of event types becomes necessary.

2. Event variables

A model of events requires, first, variables that can be used to refer to the occurrence of possible events, and to conditions on which the occurrence of events may depend. The definition of such variables must refer to situations in which events might, or might not, occur (then, in a retrospective view, one can determine which events actually did occur). The basic task therefore consists in the specification of the situations (to be used for the model). Three specifications are of particular importance.

(a) The type of the situations to be considered must be specified. For example: situations in which car drivers approach traffic lights, or situations in which people might become unemployed. We speak of a *generic situation* when reference is made to a situation, or situations, of a previously specified type.

(b) The types of the events that might occur in a generic situation must be specified. This is done by using *event variables* which will be denoted by $\dot{E}$ (or $\dot{E}_1, \dot{E}_2, \ldots$).[1] The range of an event variable is denoted by $\tilde{\mathcal{E}}$ (or $\tilde{\mathcal{E}}_1, \tilde{\mathcal{E}}_2, \ldots$) and has the general form

$$\tilde{\mathcal{E}} = \{0, 1, \ldots, m\}$$

The elements $1, \ldots, m$ refer to the possible event types; the notation $\tilde{\mathcal{E}}^* := \tilde{\mathcal{E}} \setminus \{0\}$ will be used for the subset of these values. The element 0 does not refer to a specific event type, but is sometimes necessary to express that an event did not (yet) occur.[2] For example, referring to situations where car drivers approach a traffic light, one could define an event variable $\dot{E}$ having the range $\tilde{\mathcal{E}} = \{0, 1, 2\}$ with the understanding that 1 means that the car driver stops, 2 means that the car driver does not stop, and 0 means that none of the two possible events has (yet) occurred.

It will be assumed that event types which belong to the same range (of some event variable) do exclude each other. One can define, however, two or more event variables for a generic situation so that events can occur in different combinations (examples will be discussed in § 5 and in Section 7.2). In particular, instead of an event variable $\dot{E}$ having a range $\{0, 1, \ldots, m\}$ one can always use an m-dimensional event variable $(\dot{E}_1, \ldots, \dot{E}_m)$ with the range $\{0, 1\}^m$ and components which refer to single event types: $(\dot{E}_j = 1) \equiv (\dot{E} = j)$.

(c) It must be specified whether situations should be further distinguished by properties not already part of the generic framework. Variables used to record such properties will be termed *contextual variables*. In particular, they will be called *static contextual variables* if their values cannot change during the current situation. These might be event variables which got a specific value in a previous situation, or state variables like those used in Chapter 4. On the other hand, they will be called *dynamic contextual variables* if values can change while a situation is going on.

Next, the temporal extension of generic situations must be considered. It will be assumed that a situation continues until, for the first time, an event occurs; then, as a consequence of the event, a new situation comes into being.

It follows that event variables do not always have a value that refers to a specific event. This distinguishes event variables from *state variables*. An example from Section 4.1 (§ 1) may serve as an illustration. In this example, there are three state variables: $\ddot{Y}$ records whether a bulb gives light, $\ddot{X}$ records the state of a switch, and $\ddot{Z}$ records whether a battery can provide power. All variables always have a specific value (0 or 1 in the example). An event variable, on the other hand, is always associated with a generic situation in which the variable can assume a specific value (for example, a situation in which a switch can be closed). Up to the occurrence of an event that provides a specific value, the event variable has no positive value but has the value zero, indicating that an event did not yet occur. Further, if in fact an event did occur and provide a specific value for the event variable, this value can never change.

3. Referring to event data

Events and their properties can only be determined after the events actually occurred, that is, in a retrospective perspective. This is also required in order to record events with statistical variables. In the simplest case each element ω of a reference set Ω refers to a specific situation, and there is a statistical variable

$$(X, E) : \Omega \longrightarrow \tilde{\mathcal{X}} \times \tilde{\mathcal{E}}$$

$X(\omega)$ records properties of the situation ω, and $E(\omega)$ shows the type of the event that has happened.

4. Functional event models

Instead of referring to events which actually have occurred one can think about possible events which might occur in the future (or which might have occurred in the past). This motivates the construction of models that show how the possible occurrence of events depends on identifiable conditions. In the following we consider *event models* which are defined as functional models (see Chapter 4) containing at least one event variable.

A special feature of such models is that functions with a dependent event variable always require a temporal interpretation. Such functions determine how events might occur and the corresponding event variables get their values. Consequently, any variable which functionally depends on an event variable is also an event variable. The following model can serve as an example:

$$\ddot{E}_0 \xrightarrow{f_1}\!\!\!\!\twoheadrightarrow \dot{E}_1 \xrightarrow{f_2}\!\!\!\!\twoheadrightarrow \dot{E}_2 \tag{7.1}$$

When defining the stochastic function f_1 one has to take into account that $\dot{E}_1$ can take a specific value only after a specific value of $\ddot{E}_0$ has been realized.

Therefore, the following notation will be used:

$$j_0 \longrightarrow \Pr[\dot{E}_1 \,\|\, \ddot{E}_0 = j_0]$$

The symbol $\|$ indicates that the conditional distribution refers to a situation in which the conditions (given on the right of the $\|$ symbol) are already realized facts.[3] In the same way, the stochastic function f_2 is written

$$j_1 \longrightarrow \Pr[\dot{E}_2 \,\|\, \dot{E}_1 = j_1]$$

in order to indicate that the realization of a specific value of $\dot{E}_1$ must temporally precede the realization of a specific value of $\dot{E}_2$.

5. *Models for single situations*

In the simplest case an event model refers to a generic situation specified by two variables: A deterministic contextual variable $\ddot{X}$ which records relevant properties of the situation,[4] and a stochastic event variable $\dot{E}$ specifying the events which might occur. The model then stipulates a stochastic function

$$x \longrightarrow \Pr[\dot{E} \,\|\, \ddot{X} = x] \tag{7.2}$$

defining how the probabilities of the possible events depend on values of the contextual variable $\ddot{X}$.

Several event variables for the same generic situation may be defined. The following diagram illustrates a simple model with two event variables $\dot{E}_1$ and $\dot{E}_2$:

$$\ddot{X} \twoheadrightarrow (\dot{E}_1, \dot{E}_2)$$

There is now a two-dimensional endogenous event variable $(\dot{E}_1, \dot{E}_2)$, and the corresponding stochastic function is

$$x \longrightarrow \Pr[\dot{E}_1, \dot{E}_2 \,\|\, \ddot{X} = x] \tag{7.3}$$

showing how the probability distribution of $(\dot{E}_1, \dot{E}_2)$ depends on values of $\ddot{X}$. If one assumes that both event variables refer to only one event type ($\tilde{\mathcal{E}}_1 = \tilde{\mathcal{E}}_2 = \{0, 1\}$), possible values are:

$\dot{E}_1$	$\dot{E}_2$	
1	0	the $\dot{E}_1$ event occurs first
0	1	the $\dot{E}_2$ event occurs first
1	1	both events occur simultaneously
0	0	no event occurs

The last case is impossible because we have assumed that a situation continues until at least one event occurs; $(\dot{E}_1, \dot{E}_2)$ therefore has the range $(\tilde{\mathcal{E}}_1 \times \tilde{\mathcal{E}}_2)^* = \{(1,0), (0,1), (1,1)\}$.

6. Consecutive situations

Instead of a single situation one can consider two or more consecutive situations. This allows models where event probabilities also depend on the occurrence of events in earlier situations. The school example discussed in Section 6.2 will be used for an illustration. There are two situations, σ_1 and σ_2. Depending on the parents' educational level, the decision about the type of the child's school is made in σ_1. This is captured by a contextual variable $\ddot{Z}$ with possible values 0 (low) and 1 (higher educational level) and an event variable $\dot{E}_1$ with possible values

$$\dot{E}_1 = \begin{cases} 1 \equiv \text{the child visits a school of type 1} \\ 2 \equiv \text{the child visits a school of type 2} \end{cases}$$

Then follows situation σ_2: the child visits a school of the previously decided type and an event variable $\dot{E}_2$ with possible values

$$\dot{E}_2 = \begin{cases} 1 \equiv \text{the school was not successfully completed} \\ 2 \equiv \text{the school was successfully completed} \end{cases}$$

records the educational success. The context for $\dot{E}_2$ is given by the parents' educational level and the type of the school.

There are now two possibilities to express the dependence of $\dot{E}_2$ on this context. If σ_2 is viewed as an isolated situation, the context can be represented by deterministic variables. This corresponds to version (a) in the following diagram:[5]

$$\text{(a)}\quad \begin{tikzcd} \ddot{Z} \arrow[dr, two heads] & \\ & \dot{E}_2 \\ \ddot{E}_1 \arrow[ur, two heads] & \end{tikzcd} \qquad \text{(b)}\quad \begin{tikzcd} \ddot{Z} \arrow[dr, two heads] \arrow[dd, two heads] & \\ & \dot{E}_2 \\ \dot{E}_1 \arrow[ur, two heads] & \end{tikzcd} \tag{7.4}$$

$\ddot{E}_1$ is then a deterministic variable representing the child's school type.[6] On the other hand, in version (b), $\dot{E}_1$ is an endogenous variable with the implication that the model refers to two consecutive situations. There is a stochastic function

$$z \longrightarrow \Pr[\dot{E}_1 \,\|\, \ddot{Z} = z]$$

for the situation σ_1 which determines how $\dot{E}_1$ gets its value; and there is another stochastic function

$$(z, j_1) \longrightarrow \Pr[\dot{E}_2 \,\|\, \ddot{Z} = z, \dot{E}_1 = j_1]$$

that determines, depending on the outcome of σ_1, how $\dot{E}_2$ gets a value in the consecutive situation σ_2.

7. Ramification of situations

Models may become slightly more complicated when a situation allows realizations of two or more event variables because different consecutive situations may arise depending on the realized event. To illustrate, we consider couples living in a consensual union. There are two event variables: $\dot{E}_1$ takes the value 1 (if the couple marries) or 2 (if the couple separates or one partner dies), and $\dot{E}_2$ takes the value 1 if the woman becomes pregnant. In addition there is a contextual variable $\ddot{Z}$ characterizing the situation.[7] As shown in the following diagram there are now two (or more) follow-up situations:

$$\ddot{Z} \xrightarrow{\sigma_1} (\dot{E}_1, \dot{E}_2) \begin{array}{l} \xrightarrow{\sigma_{21}} \dot{E}_1 \\ \xrightarrow[\sigma_{22}]{} (\dot{E}_2, \dot{E}_3) \end{array} \tag{7.5}$$

The stochastic function

$$z \longrightarrow \Pr[\dot{E}_1, \dot{E}_2 \,\|\, \ddot{Z} = z]$$

for situation σ_1 determines the generation of values for $\dot{E}_1$ and/or $\dot{E}_2$. If a pregnancy comes first the resulting situation is σ_{21}, and the couple can marry or separate. Correspondingly, there is a stochastic function

$$z \longrightarrow \Pr[\dot{E}_1 \,\|\, \ddot{Z} = z, \dot{E}_2 = 1]$$

On the other hand, if the first event is a realization of $\dot{E}_1$, what follows depends on which type of event occurred. The model considers a new situation only as a result of the event $\dot{E}_1 = 1$ (marriage). If this event happens the new situation is σ_{22}, and a pregnancy as well as a separation becomes possible. However, to record the possibility of a separation one cannot use the variable $\dot{E}_1$ because this variable already received a specific value ($\dot{E}_1 = 1$) in σ_1. Therefore, one needs another event variable, called $\dot{E}_3$ in the example, to express this possibility.[8] There is then a new situation σ_{22} with a corresponding stochastic function

$$z \longrightarrow \Pr[\dot{E}_2, \dot{E}_3 \,\|\, \ddot{Z} = z, \dot{E}_1 = 1]$$

A further possibility is that both event variables get a value in σ_1. The process considered by the model depicted in (7.5) is then finished.

The example shows that one needs to specify a separate (stochastic) function when a situation depends on the outcome of a previous situation.[9] The example also illustrates the two kinds of conditional distributions distinguished in § 4. $\Pr(\dot{E}_1 = 1 \,\|\, \ddot{Z} = z, \dot{E}_2 = 1)$ is the probability of a marriage in a situation that followed a previous pregnancy. On the other hand, $\Pr(\dot{E}_1 = 1 \,|\, \ddot{Z} = z, \dot{E}_2 = 1)$ refers to the situation σ_1 and provides the probability that, given a pregnancy, the couple simultaneously marries.

7.2 Event models with time axes

1. Time axes for situations

The models discussed in the previous section relate to generic situations (situations of a specific type) which provide the context for the occurrence of events (of specific predefined types). It was assumed that a situation continues until the first occurrence of an event whereby a new situation begins. This approach suffices if one is only interested in probabilities for the occurrence of events. However, often one is also interested in the duration of situations until events occur or in the dependence of event probabilities on changing conditions. This requires an explicit representation of the situation's temporal extension.

To provide this representation, a deterministic variable $\ddot{T}$ will be used that records the duration since the beginning of a situation. It will be assumed that $\ddot{T}$ has a discrete range $\mathcal{T}_0 = \{0, 1, 2, \ldots\}$ which allows one to think of a situation as a sequence of temporal locations. The beginning of a situation can, in most cases, be defined by referring to an event by which the situation comes into being. In these cases it will be assumed that the temporal location in which the event occurs equals the temporal location $\ddot{T} = 0$ of the following situation. Therefore, two successive events may occur in the same temporal location. A person, for example, can fall ill and then recover during the same day.

2. *Time-dependent event probabilities*

A time axis allows the making of event probabilities dependent on the duration since the beginning of a situation. However, a time axis is not automatically implied by event models; it must be explicitly constructed. Consider an event variable $\dot{E}$ with range $\tilde{\mathcal{E}} = \{0, 1, \ldots, m\}$ which is defined for a situation σ. If $\ddot{T}$ records the duration since the beginning of the situation, one can consider duration-dependent event probabilities

$$r_j(t) := \Pr(\dot{E} = j \parallel \ddot{T} = t) \tag{7.6}$$

The condition $\ddot{T} = t$ says that the situation has continued until the temporal location t and implies that no event occurred before t. The function $r_j(t)$ will be termed a *risk function*. It provides the probability for the occurrence of an event $\dot{E} = j$ in the temporal location t under the condition that the situation has not ended before t. Also possible is the definition of a *summarized risk function*

$$r(t) := \sum_{j \in \tilde{\mathcal{E}}^*} r_j(t) \tag{7.7}$$

which provides the probability for the occurrence of any event in temporal location t under the condition that the situation has not ended before t.

Since $\ddot{T}$ is a deterministic variable it can only be used to express conditions. However, starting from (7.6) one can also define a stochastic variable $\dot{T}$ (having the same range as $\ddot{T}$) that records the duration until the situation ends by some event. The distribution of $\dot{T}$ can be characterized by a survivor function

$$G(t) := \Pr(\dot{T} \geq t) = \prod_{k=0}^{t-1} (1 - r(k))$$

with $G(0) = \Pr(\dot{T} \geq 0) = 1$, and one gets the equation

$$r(t) = \frac{\Pr(\dot{T} = t)}{\Pr(\dot{T} \geq t)} = \frac{\Pr(\dot{T} = t)}{G(t)}$$

3. Non-occurring events

Since situations continue until at least one event occurs, models without an explicit time axis do not allow the defining of probabilities for the non-occurrence of events. This becomes possible, however, if one can view a situation as a sequence of temporal locations. If $\dot{E}$ is an event variable one can use the definition

$$\Pr(\dot{E} = 0 \,\|\, \ddot{T} = t) := 1 - r(t) \tag{7.8}$$

and interpret this as the probability that no event (of a type specified by the range of $\dot{E}$) occurs in the temporal location t.

4. Aggregation of temporal locations

Beginning with duration-dependent event probabilities, models which abstract from durations (like those discussed in Section 7.1) can be derived by aggregation over temporal locations. A simple summation

$$\Pr(\dot{E} = j) = \sum_{t=0}^{\infty} r_j(t) \Pr(\dot{T} \geq t) \tag{7.9}$$

provides the probability of the occurrence of an event $\dot{E} = j$ in the situation for which the event variable $\dot{E}$ is defined.

As an illustration consider an event variable $\dot{E}$ with the range $\tilde{\mathcal{E}} := \{0, 1, 2\}$ which is defined for a situation σ. Assuming time-constant risk functions $r_1(t) = 0.1$ and $r_2(t) = 0.2$, the following table shows how the process evolves until $t = 5$:

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

Using (7.9), summation of the time-dependent probabilities provides the overall event probabilities $\Pr(\dot{E} = 1) = 1/3$ and $\Pr(\dot{E} = 2) = 2/3$.

It is remarkable that knowing a risk function $r_j(t)$ does not suffice to calculate event probabilities $\Pr(\dot{E} = j)$. In addition one needs the survivor function $G(t)$

or, equivalently, the summarized risk function $r(t)$ defined in (7.7). It is therefore possible that an event probability $\Pr(\dot{E} = j)$ changes its value without a change in the corresponding risk function $r_j(t)$. For instance, using a risk function $r_2(t) = 0.15$ instead of $r_2(t) = 0.2$ in the example just considered would result in event probabilities $\Pr(\dot{E} = 1) = 0.4$ and $\Pr(\dot{E} = 2) = 0.6$.

5. Situations with several event variables

Aggregation over temporal locations is also possible if several event variables are defined for a situation. To illustrate we continue with the example from Section 7.1 (§ 7). In this example, there are two event variables: $\dot{E}_1$ records whether a couple living in a consensual union marries or separates, and $\dot{E}_2$ records whether the woman becomes pregnant. Introducing a time axis allows the definition of a two-dimensional risk function

$$r_{j_1 j_2}(t) = \Pr(\dot{E}_1 = j_1, \dot{E}_2 = j_2 \parallel \ddot{T} = t) \tag{7.10}$$

providing duration-dependent event probabilities. By aggregating over temporal locations one can then derive the model of Section 7.1 (§ 7) that abstracts from the temporal extension of the situation. The following table shows time-constant risk functions that can be used for a simple illustration.

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

Aggregation can be done with formula (7.9) and results in the overall event probabilities shown in the right column. They approximately equal the probabilities used for the illustration in Section 7.1.

6. Static and dynamic contextual variables

In models with duration-dependent event probabilities (risk functions) one can use static as well as dynamic contextual variables. Model specification then requires the definition of context-dependent risk functions having the general form

$$(t, z, x_t) \longrightarrow \Pr(\dot{E} = j \parallel \ddot{T} = t, \ddot{Z} = z, \ddot{X}_t = x_t)$$

In this expression $\dot{E}$ is the dependent event variable; $\ddot{Z}$ is a static contextual variable, and $\ddot{X}_t$ is a dynamic contextual variable with values that can depend on the temporal location t. The notation indicates that $\ddot{T}$ can also be considered as a dynamic contextual variable.

Event variables can be used as contextual variables when they have already assumed a specific value. Because those values cannot change, these are static contextual variables. Conditional event probabilities have the form

$$\Pr(\dot{E} = j \parallel \ddot{T} = t, \dot{E}' = j', \ldots)$$

In this expression $\dot{E}'$ is a contextual variable that has assumed the value j', meaning that an event $\dot{E}' = j'$ already has occurred in some temporal location not later than t. In any case, our interpretation of $\parallel$ implies that $\dot{E}' = j'$ is a realized condition for the possible occurrence of $\dot{E} = j$.

7. Time axes for consecutive situations

In the models so far considered the time axis is reset to zero whenever a new situation begins. Sometimes it is easier to use a single time axis $\ddot{T}$ that starts with the model's first situation and records durations since then through all subsequent situations. The following model may serve as a simple example:

$$\ddot{E}_0 \twoheadrightarrow \dot{E}_1 \twoheadrightarrow \dot{E}_2$$

As soon as some event $\ddot{E}_0 = j_0$ has taken place the first situation begins in the temporal location $\ddot{T} = 0$. Risk functions for this situation can therefore be defined as

$$r_{j_1}(t; j_0) := \Pr(\dot{E}_1 = j_1 \parallel \ddot{T} = t, \ddot{E}_0 = j_0)$$

If now an event $\dot{E}_1 = j_1$ occurs and the time axis is not reset there are two possibilities to define stochastic functions for the event variable $\dot{E}_2$. One can use conditional probabilities of the form

$$\Pr(\dot{E}_2 = j_2 \parallel \ddot{T} = t, \dot{E}_1 = j_1) \tag{7.11}$$

where the condition says that the current situation continued at least until the temporal location t and the event $\dot{E}_1 = j_1$ occurred in some temporal location before or at t. An additional notation is necessary if one wants to take into consideration when that event occurred. To allow flexible notations, we use an operator $\tau(\dot{E})$ that records the temporal location in which the event variable $\dot{E}$ got for the first time a specific value.[10] Instead of (7.11) one can then use an expression of the form

$$\Pr(\dot{E}_2 = j_2 \parallel \ddot{T} = t, \dot{E}_1 = j_1, \tau(\dot{E}_1) = t_1) \tag{7.12}$$

which, of course, implies $t \geq t_1$, and this allows the defining of a risk function

$$r_{j_2}(t; j_1, t_1) := \Pr(\dot{E}_2 = j_2 \parallel \ddot{T} = t_1 + t, \dot{E}_1 = j_1, \tau(\dot{E}_1) = t_1)$$

for $\dot{E}_2$. The definition conforms with the assumption of § 1 that whenever an event occurs in a temporal location t the following situation begins in the same temporal location.

7.3 Dynamic causality

1. Introductory remarks

The discussion of functional causality in Chapter 5 was based on a comparative notion of cause referring to two different values of a state variable. However, state variables rarely refer to events. In the current section causes will be conceived as events. We speak then of *dynamic causes* and, correspondingly, of a *dynamic conception of causality.*

This does not immediately imply a specific notion of effect. While an often meaningful idea is that events can set off processes, or modify already ongoing processes, this idea is difficult to make precise. We therefore begin with a narrower approach that again refers to events. However, in most cases it is not possible to define effects directly as events because a dynamic cause only changes probabilities for the occurrence of other events. The leading idea therefore will be that the *dynamic effect of an event e* consists in changes of subsequent event probabilities which are attributable to the occurrence of *e*.

It is evident, then, that one needs models that specify the possible events to be used for statements about effects (= changes of event probabilities). The following considerations use event models as discussed in the two previous sections. Events then consist in event variables assuming specific values.[11]

2. A definition of dynamic causality

How can one attribute specific effects to the occurrence of an event? Our approach is based on a comparison of two situations: one in which the event has occurred and another one in which the event has not occurred. The formal definition refers to a model that contains an event variable $\dot{E}_1$ (or a deterministic event variable $\ddot{E}_1$) that allows one to speak of an event $\dot{E}_1 = j_1$ (or $\ddot{E}_1 = j_1$). Possible effects can be considered with respect to all events $\dot{E}_2 = j_2$ given that also $\dot{E}_2$ belongs to the supposed model and there is a directed path leading from $\dot{E}_1$ to $\dot{E}_2$. This implies the existence of a conditional probability

$$\Pr(\dot{E}_2 = j_2 \parallel \dot{E}_1 = j_1, \ddot{Z} = z)$$

where $\ddot{Z}$ covers all contextual variables which are possibly relevant for the probability of the event $\dot{E}_2 = j_2$, and allows the following definition: The *dynamic effect* of the event $\dot{E}_1 = j_1$ with respect to the possible event $\dot{E}_2 = j_2$ in the context $\ddot{Z} = z$ is[12]

$$\Pr(\dot{E}_2 = j_2 \parallel \dot{E}_1 = j_1, \ddot{Z} = z) - \Pr(\dot{E}_2 = j_2 \parallel \dot{E}_1 = 0, \ddot{Z} = z) \tag{7.13}$$

It is assumed that the expression on the right-hand side is zero if the event $\dot{E}_2 = j_2$ is only possible after the event variable $\dot{E}_1$ has got a specific value. In the following paragraphs we use some examples for a discussion of this definition and also explore a notion of time-depending effects.

3. Illustration with a random generator

The first example is a random generator without contextual variables:

$$\ddot{E}_0 \stackrel{\sigma_1}{\twoheadrightarrow} \dot{E}_1 \stackrel{\sigma_2}{\longrightarrow} \dot{E}_2$$

The exogenous event variable $\ddot{E}_0$ with range $\tilde{\mathcal{E}}_0 = \{0, 1\}$ records whether a die has been thrown. If this happens the situation σ_1 begins in which the event variable $\dot{E}_1$ with range $\tilde{\mathcal{E}}_1 = \{0, 1, \ldots, 6\}$ records the result. The die is assumed to be unbiased so that

$$\Pr(\dot{E}_1 = j \,\|\, \ddot{E}_0 = 1) = 1/6 \quad (j = 1, \ldots, 6) \tag{7.14}$$

Then follows the situation σ_2 in which the event variable $\dot{E}_2$ with range $\tilde{\mathcal{E}}_2 = \{0, 1, 2\}$ records which type of action occurs. For simplicity it is assumed that $\dot{E}_2 = 1$ if $\dot{E}_1$ gives an uneven result and $\dot{E}_2 = 2$ if $\dot{E}_1$ gives an even result so that $\dot{E}_1$ and $\dot{E}_2$ are connected by a deterministic function.

What effects can be attributed to the events that can occur in this model? First, consider the event $\ddot{E}_0 = 1$, the throw of the die. The model provides probabilities for values of $\dot{E}_1$ and $\dot{E}_2$. To apply the definition of § 2 one also has to consider a situation in which the die is not thrown. However, in the context of the current model, no specific value for $\dot{E}_1$ is then possible, implying that $\Pr(\dot{E}_1 = j \,\|\, \ddot{E}_0 = 0) = 0$ and the dynamic effects of the event $\ddot{E}_0 = 1$ are already given by (7.14).

Now consider an endogenous event, for example $\dot{E}_1 = 1$. The model provides the conditional probabilities

$$\Pr(\dot{E}_2 = 1 \,\|\, \ddot{E}_1 = 1) = 1 \quad \text{and} \quad \Pr(\dot{E}_2 = 2 \,\|\, \ddot{E}_1 = 1) = 0$$

and again, these probabilities express the dynamic effects of $\dot{E}_1 = 1$ because, in this model, $\dot{E}_1 \neq 0$ is a necessary condition for a specific value of $\dot{E}_2$, implying that $\Pr(\dot{E}_2 = 1 \,\|\, \ddot{E}_1 = 0) = \Pr(\dot{E}_2 = 2 \,\|\, \ddot{E}_1 = 0) = 0$.

4. Two kinds of modal comparisons

Following the definition of § 2, in order to calculate the dynamic effect of an event $\dot{E} = j$ one has to compare a situation in which the event has occurred with a situation in which $\dot{E}$ has not yet any specific value. A quite different approach would be to compare a situation in which the event $\dot{E} = j$ has occurred with a situation in which another event, say $\dot{E} = j'$, has occurred.

The school example discussed in Section 7.1 (§ 6) can serve to illustrate the differences. As possible causes one can consider the events $\dot{E}_1 = 1$ (the child visits a school of type 1) and $\dot{E}_1 = 2$ (the child visits a school of type 2). Following the definition of § 2 one can attribute dynamic effects with respect to possible values of $\dot{E}_2$ to each of these events separately; and because $\dot{E}_1$ must assume some specific value before any event $\dot{E}_2 = j_2$ can occur, these dynamic effects are directly given by the context-dependent probabilities

$$\Pr(\dot{E}_2 = j_2 \parallel \dot{E}_1 = 1, \ddot{Z} = z) \quad \text{and} \quad \Pr(\dot{E}_2 = j_2 \parallel \dot{E}_1 = 2, \ddot{Z} = z)$$

Of course, one can afterwards compare the effects of $\dot{E}_1 = 1$ and $\dot{E}_1 = 2$; but this is not required for a calculation of the effects that can be attributed to both events separately.

It is noteworthy that the kind of comparison which is used for the definition of dynamic causality in § 2 is possible only with event variables, not with state variables. Because state variables always assume some specific value one can only compare implications of different values (as was done in the definition of effects of comparative causes in Chapter 5).

5. *Exogenous intervening causes*

So far we have discussed examples where an event variable $\dot{E}_1$ (specifying a cause) must assume some specific value in order that an event $\dot{E}_2 = j_2$, specifying an effect, can occur. We speak of *intervening causes* if this is not the case. The following model provides a simple illustration.

$$\begin{tikzcd} \ddot{E}_0 \arrow[r, twoheadrightarrow] & \dot{E}_2 \\ \ddot{E}_1 \arrow[ur, twoheadrightarrow] & \end{tikzcd} \tag{7.15}$$

The event $\ddot{E}_0 = 1$ means that a person falls ill. Subsequently the person can recover ($\dot{E}_2 = 1$) or die ($\dot{E}_2 = 2$). Also possible is an event $\ddot{E}_1 = 1$ that consists in the application of some therapy. The difference from the examples considered in previous paragraphs is that events $\dot{E}_2 = j_2$ do not require that $\ddot{E}_1$ has already assumed some specific value. Therefore two conditional distributions are necessary for a full specification:

$$\Pr(\dot{E}_2 = j_2 \parallel \ddot{E}_0 = 1, \ddot{E}_1 = 1) \quad \text{and} \quad \Pr(\dot{E}_2 = j_2 \parallel \ddot{E}_0 = 1, \ddot{E}_1 = 0)$$

Effects of the event $\ddot{E}_1 = 1$ with respect to the possible events $\dot{E}_2 = 1$ and $\dot{E}_2 = 2$ can then be calculated from these probabilities.

6. *Endogenous intervening causes*

We now consider the possibility of defining intervening causes by an endogenous event variable. To illustrate the discussion we take up the example of Section 7.1 (§ 7) which refers to consensual unions. To connect to the discussion in the previous paragraph the following notation is used: A consensual union begins with the event $\ddot{E}_0 = 1$, $\dot{E}_2$ records whether the couple marries ($\dot{E}_2 = 1$) or separates ($\dot{E}_2 = 2$), and $\dot{E}_1 = 1$ records the occurrence of a pregnancy. The question concerns the effect of a pregnancy on the probability of a marriage. $\dot{E}_1 = 1$ is obviously an intervening cause (if at all).

Since $\dot{E}_1$ is now an endogenous event variable, the beginning of a consensual union ($\ddot{E}_0 = 1$) implies some probability for the occurrence of an event $\dot{E}_1$ (to be considered as an intervening cause). To provide a graphical picture one could replace $\ddot{E}_1$ by $\dot{E}_1$ in (7.15) and add an arrow from $\ddot{E}_0$ to $\dot{E}_1$. However, the diagram would be misleading because it would obscure the fact that, after forming the consensual union, both event variables, $\dot{E}_1$ and $\dot{E}_2$, can get specific values. Preferable is therefore a graphical illustration like (7.5) in Section 7.1. For the current discussion a simplified version suffices:

$$\ddot{E}_0 \xrightarrow{\sigma_1}\!\!\!\!\rightarrow (\dot{E}_1, \dot{E}_2) \xrightarrow{\sigma_2}\!\!\!\!\rightarrow \dot{E}_2 \tag{7.16}$$

Corresponding to situation σ_1 there are conditional probabilities

$$\Pr(\dot{E}_1 = j_1, \dot{E}_2 = j_2 \,\|\, \ddot{E}_0 = 1)$$

from which probabilities of a marriage and a separation before, or simultaneous with, the occurrence of a pregnancy, follow:

$$\begin{aligned}&\Pr(\dot{E}_2 = j_2 \,\|\, \ddot{E}_0 = 1) = \\ &\qquad \Pr(\dot{E}_1 = 0, \dot{E}_2 = j_2 \,\|\, \ddot{E}_0 = 1) + \Pr(\dot{E}_1 = 1, \dot{E}_2 = j_2 \,\|\, \ddot{E}_0 = 1)\end{aligned}$$

Relevant for the causal question is a subsequent situation σ_2 that follows when, in the first situation, a pregnancy, but neither a marriage nor a separation occurs. In this situation either a marriage or a separation occurs, with corresponding probabilities given by

$$\Pr(\dot{E}_2 = j_2 \,\|\, \ddot{E}_0 = 1, \dot{E}_1 = 1)$$

Dynamic effects of a pregnancy can finally be calculated by the difference

$$\Pr(\dot{E}_2 = j_2 \,\|\, \ddot{E}_0 = 1, \dot{E}_1 = 1) - \Pr(\dot{E}_2 = j_2 \,\|\, \ddot{E}_0 = 1)$$

Using the numerical values from Section 7.1 (§ 7) one finds that a pregnancy increases the probability of a marriage by $0.6 - 0.33 = 0.27$ and decreases the probability of a separation by $0.4 - 0.51 = -0.11$.

7. Concurrent causes

Concurrent causes can occur if the probability of an event depends on two (or more) different event variables. The following example is often discussed in the philosophical literature:

> Fred goes into the desert. He has two enemies, Zack and Mack: Zack aims to poison him by putting poison in his water-bottle; Mack aims to cause him to die of dehydration by punching a hole in his bottle.
>
> (Barker 2004: 124)

In order to set up a model, the following event variables are used: $\ddot{E}_1 = 1$ if Mack punches a hole in the water-bottle, $\ddot{E}_2 = 1$ if Zack puts poison in the water-bottle, $\dot{E}_2' = 1$ if Fred drinks the poisoned water, $\dot{E}_3 = 1$ if Fred dies. The functional relationships are assumed as follows:

$$
\begin{tikzcd}
\ddot{E}_1 \arrow[rr, two heads] \arrow[dr] & & \dot{E}_3 \\
\ddot{E}_2 \arrow[r, two heads] & \dot{E}_2' \arrow[ur] &
\end{tikzcd}
\tag{7.17}
$$

The model allows the consideration of several causal questions. In any case, one needs the conditional probabilities

$$
\begin{aligned}
&\Pr(\dot{E}_3 = 1 \,|\, \ddot{E}_1 = j_1, \ddot{E}_2 = j_2) = \\
&\quad \Pr(\dot{E}_3 = 1 \,|\, \ddot{E}_1 = j_1, \dot{E}_2' = 0)\Pr(\dot{E}_2' = 0 \,|\, \ddot{E}_1 = j_1, \ddot{E}_2 = j_2) + \\
&\quad \Pr(\dot{E}_3 = 1 \,|\, \ddot{E}_1 = j_1, \dot{E}_2' = 1)\Pr(\dot{E}_2' = 1 \,|\, \ddot{E}_1 = j_1, \ddot{E}_2 = j_2)
\end{aligned}
$$

To illustrate, the following values are assumed:

j_1	j_2	$\mathrm{E}(\dot{E}_2' \,\|\, \ddot{E}_1 = j_1, \ddot{E}_2 = j_2)$
0	0	0.0
0	1	0.9
1	0	0.0
1	1	0.8

j_1	j_2'	$\mathrm{E}(\dot{E}_3 \,\|\, \ddot{E}_1 = j_1, \dot{E}_2' = j_2')$
0	0	0.0
0	1	0.90
1	0	0.85
1	1	0.95

One then finds: If $\ddot{E}_2 = 0$, the occurrence of $\ddot{E}_1 = 1$ raises the probability of Fred's death from zero to 0.85; and if $\ddot{E}_1 = 0$, the occurrence of $\ddot{E}_2 = 1$ raises the probability of Fred's death from zero to 0.81. The philosophical discussion mainly concerns a situation where both events occur simultaneously. There are then different possibilities to describe effects. One can consider a joint effect: $\mathrm{E}(\dot{E}_3 \,|\, \ddot{E}_1 = 1, \ddot{E}_2 = 1) = 0.93$. Alternatively, one can consider each event as a covariate context for the other event resulting in

$$
\mathrm{E}(\dot{E}_3 \,|\, \ddot{E}_1 = 1, \ddot{E}_2 = 1) - \mathrm{E}(\dot{E}_3 \,|\, \ddot{E}_1 = 0, \ddot{E}_2 = 1) = 0.93 - 0.89 = 0.04
$$

$$
\mathrm{E}(\dot{E}_3 \,|\, \ddot{E}_1 = 1, \ddot{E}_2 = 1) - \mathrm{E}(\dot{E}_3 \,|\, \ddot{E}_1 = 1, \ddot{E}_2 = 0) = 0.93 - 0.85 = 0.08
$$

In both cases, the events contribute positively to the probability of Fred's death. This depends, however, on the chosen parameters. For example, assuming deterministic preemption ($\mathrm{E}(\dot{E}_2' \,|\, \ddot{E}_1 = 1, \ddot{E}_2 = 1) = 0$ instead of 0.8), the contribution of $\ddot{E}_2 = 1$ would disappear:

$$\mathrm{E}(\dot{E}_3 \,|\, \ddot{E}_1 = 1, \ddot{E}_2 = 1) - \mathrm{E}(\dot{E}_3 \,|\, \ddot{E}_1 = 1, \ddot{E}_2 = 0) = 0.85 - 0.85 = 0$$

8. Time-dependent dynamic effects

If an event model contains a time axis one can consider time-dependent dynamic effects. In the simplest case there are two event variables, $\dot{E}_1$ and $\dot{E}_2$, and $\dot{E}_2$ can assume a specific value only if some event $\dot{E}_1 = j_1$ has occurred earlier. This implies that $\Pr(\dot{E}_2 = j_2 \parallel \dot{E}_1 = 0) = 0$, and time-dependent effects of an event $\dot{E}_1 = j_1$ can directly be expressed by a risk function

$$\Pr(\dot{E}_2 = j_2 \parallel \dot{E}_1 = j_1, \ddot{T} = t)$$

where it is assumed that the time axis $\ddot{T}$ begins with the occurrence of the event $\dot{E}_1 = j_1$.

Intervening causes require a slightly different approach. We begin with the model discussed in § 5. If the exogenous cause $\ddot{E}_1 = 1$ occurs in a temporal location $\tau(\ddot{E}_1) = t_1$ one can first consider a risk function

$$r_{j_2}(t; 1, t_1) := \Pr(\dot{E}_2 = j_2 \parallel \ddot{E}_0 = 1, \ddot{E}_1 = 1, \tau(\ddot{E}_1) = t_1, \ddot{T} = t_1 + t) \tag{7.18}$$

that begins at t_1. A sensible comparison arises from assuming that an event $\ddot{E}_1 = 1$ did not occur until, and including, the temporal location t_1 which allows consideration of a risk function

$$r_{j_2}(t; 0, t_1) := \Pr(\dot{E}_2 = j_2 \parallel \ddot{E}_0 = 1, \ddot{E}_1 = 0, \ddot{T} = t_1 + t) \tag{7.19}$$

The time-dependent effect of an event $\ddot{E}_1 = 1$ that occurs in t_1 can be defined, then, by the difference of the risk functions (7.18) and (7.19):

$$\text{Time-dependent effect} = r_{j_2}(t; 1, t_1) - r_{j_2}(t; 0, t_1) \tag{7.20}$$

It is thus possible that the effect not only depends on the duration since the occurrence of the cause, but also on the temporal location t_1, that is the duration between the beginning of the process, defined by $\ddot{E}_0 = 1$, and the occurrence of the cause $\ddot{E}_1 = 1$.

Since the risk functions in (7.18) and (7.19) can also be defined if an endogenous event variable $\dot{E}_1$ is used instead of $\ddot{E}_1$, the same approach can be used to express time-dependent effects of endogenous intervening causes.

9. Local and integrated effects

The definition (7.20) relates to risk functions and therefore provides effects which are temporally local. A further question is how an intervening cause changes the overall probability (integrated over all subsequent temporal locations) for the occurrence of another event. An answer can be derived analogously to the reasoning in Section 7.2 (§ 4).

As an illustration consider again the model (7.15). In order to calculate the integrated effect of a cause $\ddot{E}_1 = 1$ that occurred in a temporal location t_1 one needs the risk functions defined in (7.18) and (7.19) and also probabilities for the continuation of the situation that began in t_1. Since these probabilities also depend on whether the cause event occurred, two different duration variables and distributions are required: A duration variable $\dot{T}_1$ with distribution

$$\Pr(\dot{T}_1 \geq t_1) = \prod_{k=0}^{t-1} (1 - r(k; 1, t_1)$$

where $r(k; 1, t_1) := \sum_j r_j(k; 1, t_1)$, records the situation's duration if the cause event $\ddot{E}_1 = 1$ occurred in t_1; and another duration variable $\dot{T}_0$ with an analogously defined distribution records the situation's duration if the cause event did not occur in t_1. An integrated effect can then be calculated as the difference

$$\sum_{t=0}^{\infty} r_{j_2}(t; 1, t_1) \Pr(\dot{T}_1 \geq t) - r_{j_2}(t; 0, t_1) \Pr(\dot{T}_0 \geq t) \qquad (7.21)$$

To illustrate we use the constant risk functions of the example in Section 7.2 (§ 4): $r_1(t; 0, t_1) = 0.1$ and $r_2(t; 0, t_1) = 0.2$. This implies that the integrated effect does not depend on the temporal location in which the cause event occurred, so one can simply assume that due to the occurrence of an event $\ddot{E}_1 = 1$, the risk functions change to $r_1(t; 1, t_1) = 0.15$ and $r_2(t; 0, t_1) = 0.25$. The integrated effect of the event $\ddot{E}_1 = 1$ is then

for $\dot{E}_2 = 1$: $\quad 0.375 - 1/3 \approx \;\; 0.042$

for $\dot{E}_2 = 2$: $\quad 0.625 - 2/3 \approx -0.042$

The example shows again that no simple relationship exists between risk functions and temporally integrated event probabilities.

8 Multilevel and population-level models

8.1 Conceptual frameworks
1. Deterministic population-level models
2. Populations without identifiable units
3. Distribution-dependent regression functions
4. Corresponding individual-level models?
5. A version of multilevel models
6. Stochastic population-level models
7. Deriving multilevel models
8. Endogenous population-level variables

8.2 Models of statistical processes
1. Process frames and models
2. Simple diffusion models
3. Pure individual-level models
4. Modeling interdependencies

8.3 Functional causality and levels
1. Distribution-dependent causation
2. Effects of changing distributions
3. Population-level effects
4. Relationships between levels
5. Self selection with constraints
6. Time-dependent effects
7. Mixing group-level effects

Functional models can be constructed for any kind of object or situation as far as their properties can be represented by variables. An interesting distinction arises if one considers a collection of individual units, say $\Omega = \{\omega_1, \ldots, \omega_n\}$. Functional models may then refer either to individual members of Ω, or to Ω as a set. In the former case, models connect variables characterizing individual members of Ω; such models will therefore be called *individual-level models*. In the latter case models connect statistical variables or distributions that characterize Ω as a set of individual units; such models will be called *population-level models*.

The two types of models correspond to different questions. The school example introduced in Section 4.3 provides an illustration. On the one hand, one can refer to an individual child and ask how his or her educational success depends on the school type and the parents' educational level. This is the question of an individual-level model. On the other hand, one can refer to a population of children and ask in which way the proportion of successful children depends on the distribution of school types and parents' educational levels. This would be the question of a population-level model.

Statistical social research is predominantly concerned with individual-level models.[1] It is important to recognize, however, that answers to questions concerning connections between statistical distributions cannot, in general, be answered from individual-level models. They require population-level models which allow taking into account constraints and interdependencies at the population level. In fact, this may already be necessary for questions relating to the individual level. It will often be the case that what happens on the individual level depends on circumstances to be defined on a population level. Individual-level models must then be extended into (some version of) multilevel models.

Collections of individual units can be small (e.g. classes), or of intermediate size (e.g. neighborhoods and communities), or represent whole countries. In this chapter, I indiscriminately speak of populations and do not consider distinctions relating to their size. The first section introduces conceptual frameworks for deterministic and stochastic population-level models. These models relate to statistical populations consisting of individual units which cannot be identified and therefore differ from models for structured units (as defined in Section 2.2). The first section also introduces a version of multilevel models which combine individual-level and population-level variables. The second section briefly discusses models of statistical processes using diffusion models for an illustration. The third section takes up the notion of functional causality, introduced in Chapter 5, and discusses how this notion can be used for multilevel and population-level models.

8.1 Conceptual frameworks

1. Deterministic population-level models

Suppose that a deterministic individual-level model, say $\ddot{X} \longrightarrow \ddot{Y}$, is applicable to all members of a fixed reference set $\Omega = \{\omega_1, \ldots, \omega_n\}$. The ranges of $\ddot{X}$ and $\ddot{Y}$ will be denoted by $\tilde{\mathcal{X}}$ and $\tilde{\mathcal{Y}}$, respectively.

In a first step one can arbitrarily create values of $\ddot{X}$ for each $\omega \in \Omega$ and thereby create values of a statistical variable $X : \Omega \longrightarrow \tilde{\mathcal{X}}$. This implies a statistical distribution $\mathrm{P}[X]$. Given then, for each member of Ω, a value of $\ddot{X}$, one can use the individual-level model to derive corresponding values of $\ddot{Y}$ and thereby create values of a statistical variable $Y : \Omega \longrightarrow \tilde{\mathcal{Y}}$ that implies a distribution $\mathrm{P}[Y]$. In this way, the individual-level model for the members of Ω can be used to construct a functional model for Ω that associates with each distribution $\mathrm{P}[X]$ another distribution $\mathrm{P}[Y]$.

Notice that the derived model for Ω connects modal variables, but variables of a special kind, having statistical distributions as values. Such variables will be called *deterministic population-level variables* and denoted by $\ddot{X}^*$, $\ddot{Y}^*$, and so on. Ranges of possible values will be denoted, respectively, by $\mathcal{D}[\Omega, \tilde{\mathcal{X}}]$, $\mathcal{D}[\Omega, \tilde{\mathcal{Y}}]$, and so on. For example, $\mathcal{D}[\Omega, \tilde{\mathcal{X}}]$ is the set of all distributions of variables $X : \Omega \longrightarrow \tilde{\mathcal{X}}$ which can be defined for Ω with the property space $\tilde{\mathcal{X}}$.[2] Of course, population-level variables can also refer to quantities derived from statistical distributions (e.g. the number of units exhibiting some specified property), and ranges will then be different from the standard form.

Using these notations, the model just considered can be depicted by a diagram of the following form:

$$\ddot{X}^* \longrightarrow \ddot{Y}^* \tag{8.1}$$

where the arrow $\longrightarrow$ refers to a deterministic function which assigns to each possible value of $\ddot{X}^*$ exactly one value of $\ddot{Y}^*$. It is a functional model for Ω that shows how values of the endogenous variable $\ddot{Y}^*$ depend on values of the exogenous variable $\ddot{X}^*$. So it is an example of a *deterministic population-level model*.

2. Populations without identifiable units

In accordance with the statistical approach it will be assumed that population-level models concern populations whose individual members cannot be identified. Values of population-level variables are therefore taken as statistical distributions (or quantities derived from such distributions). When referring to the statistical variables from which the distributions are derived, one must bear in mind that population-level models do not distinguish between statistical variables having the same distribution.

In order to emphasize the statistical approach in which individual units are not identifiable, we avoid using vector-valued population variables having the form $(\ddot{X}_1, \ldots, \ddot{X}_n)$ where the components refer to the individual members of Ω. This notation is only useful if the purpose is to develop models for structured units (as defined in Section 2.2). Moreover, it must be kept in mind that the population-level models considered in the present chapter do not refer to structured units but to statistical populations, and hence that assumptions about relational structures must not presuppose that individual units can be identified.

3. Distribution-dependent regression functions

The model in (8.1) connects two marginal distributions. In order to use conditional distributions the model must be extended to

$$\ddot{X}^* \longrightarrow (\ddot{X}, \ddot{Y})^*$$

that associates with each distribution $\mathrm{P}[X]$ a two-dimensional distribution $\mathrm{P}[X, Y]$ having $\mathrm{P}[X]$ as a marginal distribution. Only in the extended formulation can a regression function

$$x \longrightarrow \mathrm{P}[Y \mid X = x]$$

be formulated. However, this regression function may also depend on $\mathrm{P}[X]$ (to be distinguished from the value x). It will be called, then, a *distribution-dependent regression function* and written as

$$(x, \mathrm{P}[X]) \longrightarrow \mathrm{P}[Y \mid X = x, \ddot{X}^* = \mathrm{P}[X]] \tag{8.2}$$

As an example, think of a population, Ω, consisting of n persons who want to use a train. The train has two classes ($j = 1, 2$), and s_j is the number of seats in class j. Variables are defined as $X(\omega) = j$ if ω buys a ticket for class j, and $Y(\omega) = 1$ if ω gets a seat (in the class for which he or she bought a ticket), otherwise $Y(\omega) = 0$. Due to the constraint that results from the limited number of seats, the regression function will be dependent on the distribution of X. As an example one can think of a function

$$\mathrm{P}(Y = 1 \mid X = j, \ddot{X}^* = \mathrm{P}[X]) = \min\{1, s_j/(n\,\mathrm{P}(X = j))\} \tag{8.3}$$

4. Corresponding individual-level models?

If a population-level model results from independent repetitions of an individual-level model, as was assumed in § 1, it obviously leads to distribution-independent regression functions. On the other hand, starting from a population-level model that implies distribution-dependent regression functions, it is impossible to specify a corresponding pure individual-level model. The reason is simply that it is then necessary to refer to a statistical distribution, and this requires a population-level variable.

The example of the previous paragraph can serve to illustrate the argument. It is quite possible to define an individual-level variable, $\ddot{X}$, that records whether a person buys a first or second class ticket. Similarly, it is possible to define an individual-level dependent variable. This obviously must be a stochastic variable, say $\dot{Y}$, with $\dot{Y} = 1$ if the person gets a seat, and otherwise $\dot{Y} = 0$. However, there is no function $x \longrightarrow \Pr[\dot{Y} \mid \ddot{X} = x]$ because the relationship between $\ddot{X}$ and $\dot{Y}$ depends on the distribution of a *statistical* variable, X, that provides the distribution of tickets bought in the relevant population.

5. A version of multilevel models

A stochastic version of distribution-dependent regression functions (as exemplified by (8.2)) has the following form:

$$(x, \mathrm{P}[X]) \longrightarrow \Pr[\dot{Y} \mid \ddot{X} = x, \ddot{X}^* = \mathrm{P}[X]] \tag{8.4}$$

The corresponding functional model may be depicted as

$$\begin{tikzcd}[row sep=small] \ddot{X} \arrow[dr, two heads] & \\ & \dot{Y} \\ \ddot{X}^* \arrow[ur, two heads] & \end{tikzcd} \tag{8.5}$$

showing that the probability distribution of the endogenous variable depends not only on an individual-level variable, $\ddot{X}$, but also on a population-level variable, $\ddot{X}^*$.

A model of this kind can be called a *multilevel model* since it combines individual-level and population-level variables. Corresponding to (8.3), one would get the formulation

$$\Pr(\dot{Y} = 1 \,|\, \ddot{X} = j, \ddot{X}^* = \mathrm{P}[X]) = \min\{1, s_j/(n\,\mathrm{P}(X = j))\} \tag{8.6}$$

In contrast to the deterministic population-level model (8.3), this model implies some kind of individual-level stochastic process in which probabilities for getting a seat are defined for generic individuals. Multilevel models having the form (8.5) are therefore different from population-level models. Their focus on an individual-level endogenous variable suggests thinking of these models as a version of individual-level models that incorporate at least one exogenous population-level variable. In general, it is not required that the model also includes a corresponding individual-level variable. A model that contains both variables, say $\ddot{X}^*$ and $\ddot{X}$, as assumed in the example, obviously implies restrictions on the idea of independent repetitions. The assignment of values to $\ddot{X}$ must be consistent with the specified distribution $\ddot{X}^* = \mathrm{P}[X]$, and must be completed before a value of $\dot{Y}$ can be generated. Note that, according to (8.6), the number of persons assigned to seats may well exceed the number of available seats. (This will be further discussed in § 8.)

6. Stochastic population-level models

In § 1 a deterministic population-level model was derived from independent repetitions of a deterministic individual-level model. Instead one can start from a stochastic model, say $\ddot{X} \twoheadrightarrow \dot{Y}$. As an example think of Ω as a set of persons. Values of $\dot{Y}$ record their educational level, and values of $\ddot{X}$ record the educational level of their parents. Both variables are binary, 0 represents a "low" and 1 represents a "high" educational level. The stochastic function is given by

$$\Pr(\dot{Y} = 1 \,|\, \ddot{X} = 0) = \pi_{01} \quad \text{and} \quad \Pr(\dot{Y} = 1 \,|\, \ddot{X} = 1) = \pi_{11}$$

Now let $\mathrm{P}[X] \in \mathcal{D}[\Omega, \tilde{\mathcal{X}}]$ be the distribution of a statistical variable X, representing the parents' educational levels. The number of persons whose parents have a low or a high educational level are then, respectively, $n_0 := n\,\mathrm{P}(X = 0)$ and $n_1 := n\,\mathrm{P}(X = 1)$. Now, given the distribution $\mathrm{P}[X]$, what can be said about the distribution of the educational levels of the members of Ω? Since the individual-level model is stochastic, it is not possible to derive a unique distribution. Instead one has to consider a probability distribution for these distributions, that is, a probability distribution for the elements of $\mathcal{D}[\Omega, \tilde{\mathcal{Y}}]$.

In the present example one can use a stochastic variable, say $\dot{K}^*$, that records the number of persons in Ω having a high educational level. Possible values are $k = 0, \ldots, n$. Analogously defined are variables $\dot{K}_0^*$ and $\dot{K}_1^*$ referring, respectively, to the subgroups of persons whose parents have a low or a high educational level. Since values result from independent repetitions, the distributions are given by

$$\Pr(\dot{K}_j^* = k) = \binom{n_j}{k} \pi_{j1}^k (1 - \pi_{j1})^{n_j - k}$$

(for $j = 0, 1$), and one gets the mixture distribution

$$\Pr(\dot{K}^* = k) = \sum_{l=\max\{0,k-n_1\}}^{\min\{k,n_0\}} \Pr(\dot{K}_0^* = l) \Pr(\dot{K}_1^* = k - l)$$

for $\dot{K}^* = \dot{K}_0^* + \dot{K}_1^*$. Notice that $\dot{K}^*$ is not an individual-level variable but refers to the population. $\Pr(\dot{K}^* = k)$ is the probability of a distribution of educational levels in Ω:

$$\Pr(\dot{K}^* = k) \; = \; \text{Probability of } (\mathrm{P}(Y = 1) = k/n)$$

In contrast to the deterministic population-level variables introduced in § 1, the variable $\dot{K}^*$ provides an example of a *stochastic population-level variable*. The '*' sign is used to distinguish these variables from corresponding individual-level variables. If not otherwise suggested by the application context, ranges will be taken as sets of statistical distributions. For example, one might use a stochastic population-level variable $\dot{Y}^*$ having the range $\mathcal{D}[\Omega, \tilde{\mathcal{Y}}]$. There is then, for each $\mathrm{P}[Y] \in \mathcal{D}[\Omega, \tilde{\mathcal{Y}}]$, a (conditional) probability $\Pr(\dot{Y}^* = \mathrm{P}[Y])$, to be interpreted as the probability of the statistical distribution $\mathrm{P}[Y]$.

A *stochastic population-level model* will be defined as a functional model that has at least one endogenous stochastic population-level variable. In the simplest case (as illustrated by the example) the model can be depicted as

$$\ddot{X}^* \longrightarrow\!\!\!\!\rightarrow \dot{Y}^* \tag{8.7}$$

The model connects a deterministic exogenous variable $\ddot{X}^*$ with a stochastic endogenous variable $\dot{Y}^*$. The functional relationship is stochastic (depicted by the arrow $\longrightarrow\!\!\!\!\rightarrow$) and may be written explicitly as

$$\mathrm{P}[X] \longrightarrow \Pr(\dot{Y}^* = \mathrm{P}[Y] \,|\, \ddot{X}^* = \mathrm{P}[X]) \tag{8.8}$$

To each value $\mathrm{P}[X]$ of $\ddot{X}^*$ the function assigns a probability distribution for the possible values of $\dot{Y}^*$.

7. Deriving multilevel models

The marginal model (8.7) suffices to derive population-level regression functions. The derivation of a multilevel model that explicitly also refers to the individual level requires a population-level model $\ddot{X}^* \longrightarrow\!\!\!\!\rightarrow (\ddot{X}, \dot{Y})^*$ in which, conditional on a given distribution P[X], probabilities for the joint distributions P[X, Y] are defined. The stochastic function may be written as

$$\mathrm{P}[X] \longrightarrow \Pr((\ddot{X}, \dot{Y})^* = \mathrm{P}[X, Y] \,|\, \ddot{X}^* = \mathrm{P}[X])$$

and is defined for all distributions $\mathrm{P}[X, Y] \in \mathcal{D}[\Omega, \tilde{\mathcal{X}} \times \tilde{\mathcal{Y}}]$ having a fixed marginal distribution P[X].

This then allows one to derive a multilevel model in the sense of § 5. The model includes a deterministic population-level variable $\ddot{X}^*$, a correspondingly defined deterministic individual-level variable $\ddot{X}$, and a stochastic individual-level variable $\dot{Y}$. The distribution of $\dot{Y}$ is defined by

$$\Pr(\dot{Y} = y \,|\, \ddot{X} = x, \ddot{X}^* = \mathrm{P}[X]) :=$$

$$\sum_{\mathrm{P}[X,Y] \in \mathcal{D}[\Omega, \tilde{\mathcal{X}} \times \tilde{\mathcal{Y}}]} \mathrm{P}(Y = y \,|\, X = x) \Pr((\ddot{X}, \dot{Y})^* = \mathrm{P}[X, Y] \,|\, \ddot{X}^* = \mathrm{P}[X])$$

with the understanding that the summation only includes joint distributions with a given marginal distribution. If there is no distribution-dependence, as in the example of § 6, the expression reduces to a simple individual-level model providing values of $\Pr(\dot{Y} = y \,|\, \ddot{X} = x)$. In general, however, one needs a multilevel model that allows one to take into account distribution-dependent relationships.

8. Endogenous population-level variables

An important task of functional models is to provide explicit representations of the (substantial) processes that generate the outcomes of interest. Multilevel models are particularly interesting because they allow one to investigate how individual outcomes may also depend on endogenously generated values of population-level variables (statistical distributions). To illustrate, assume that all of the n children in the population Ω want to attend a school that has capacity for s children. Selection depends on an admission test, and the probability of successfully passing the test depends on the parents' educational level recorded by X (0 low, 1 high). The probabilities are, respectively, π_0 and π_1. Finally, if the number of successful children is not greater than s, all of them are admitted; otherwise s of them are randomly selected.

In order to set up a multilevel model the following variables will be used. $\ddot{X}^*$ provides the distribution of parents' educational levels in the population; $\ddot{X}$ is the corresponding individual-level variable. $\dot{K}$ records whether a generic individual successfully passes the admission test (1 if successful, 0 otherwise); the corresponding population-level variable $\dot{K}^*$ records the number of individuals who successfully pass the test. $\dot{Y}=1$ if an individual is accepted to visit the school, otherwise $\dot{Y}=0$. The model may then be depicted as follows.

$$\begin{array}{ccccc} \ddot{X} & \longrightarrow\!\!\!\!\!\!\rightarrow & \dot{K} & \searrow & \\ & & & & \dot{Y} \\ \ddot{X}^* & \longrightarrow\!\!\!\!\!\!\rightarrow & \dot{K}^* & \nearrow & \end{array} \tag{8.9}$$

Three processes can be distinguished.

(a) The individual-level process $\ddot{X} \twoheadrightarrow \dot{K}$ which, by assumption, is not distribution-dependent and can be independently repeated. The conditional probabilities are simply given by $\Pr(\dot{K}=1\,|\,\ddot{X}=j)=\pi_j$.
(b) The population-level process $\ddot{X}^* \twoheadrightarrow \dot{K}^*$ which, of course, depends on the distribution P[X] given as value of $\ddot{X}^*$ but results from independent repetitions of the individual-level process mentioned in (a). Calculation of $\Pr(\dot{K}^*=k\,|\,\ddot{X}^*=\mathrm{P}[X])$ can be done as illustrated in § 6.
(c) Finally there is a process that generates values of $\dot{Y}$ depending on $\dot{K}$ and $\dot{K}^*$. It might be considered as an individual-level process (as suggested in § 5). However, examples of this process cannot be realized in the form of independent repetitions of the individual-level model.

Given a value of $\dot{K}^*$, a distribution of the population-level variable $\dot{Y}^*$ that records the number of finally accepted individuals, cannot be generated from independent repetitions. In fact, in the example a deterministic function connects $\dot{K}^*$ with $\dot{Y}^*$:

$$\Pr(\dot{Y}^*=y\,|\,\dot{K}^*=k)=\begin{cases}1 & \text{if } y=\min\{k,s\}\\ 0 & \text{otherwise}\end{cases}$$

A reduced individual-level model can further illustrate the limitations and possibilities of independent repetitions (due to constraints on endogenous population-level variables). One can start from

$$\Pr(\dot{Y}=1\,|\,\ddot{X}=j,\ddot{X}^*=\mathrm{P}[X],\dot{K}^*=k)=\pi_j \min\left\{1,\frac{s}{k}\right\}$$

Assuming that $\dot{K}^*$, given $\ddot{X}^*$, is (approximately) independent of $\ddot{X}$, it is possible to average over possible outcomes:

$$\Pr(\dot{Y}=1\,|\,\ddot{X}=j,\ddot{X}^*=\mathrm{P}[X])= \tag{8.10}$$
$$\pi_j \sum_{k=0}^{n} \min\left\{1,\frac{s}{k}\right\} \Pr(\dot{K}^*=k\,|\,\ddot{X}^*=\mathrm{P}[X])$$

The sum on the right-hand side can be interpreted as the mean proportion of finally accepted children, measured as part of the children who successfully passed the test, the mean taken over all possible repetitions of the process. Assuming $n=100$,

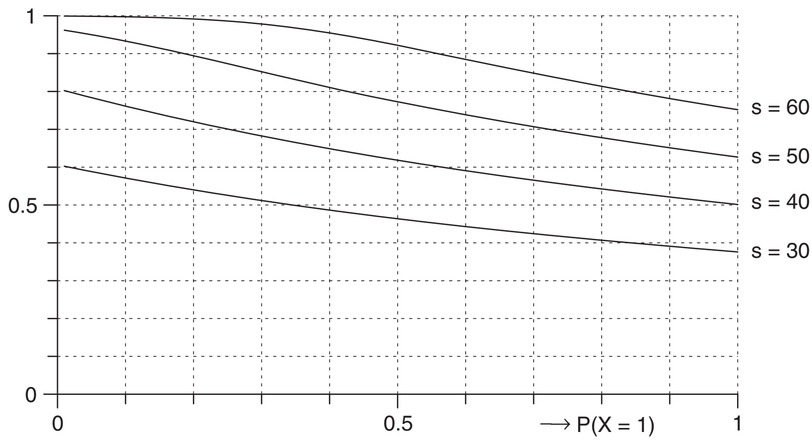


Figure 8.1 Dependence of the sum term in (8.10) on $P(X = 1)$ for different values of s. Parameters: $n = 100$, $\pi_0 = 0.5$, and $\pi_1 = 0.8$.

$\pi_0 = 0.5$, and $\pi_1 = 0.8$, Figure 8.1 illustrates the dependence of this mean proportion on $P(X = 1)$ for different values of s. Multiplication with π_j would show how the probability of being accepted depends on the distribution of X.

Thus, (8.10) can well be used to derive expectations of individual outcomes ("*ceteris paribus*"). However, it cannot be used to generate a distribution of outcomes in the population because independent repetitions will not guarantee an adherence to the constraint s.

8.2 Models of statistical processes

1. Process frames and models

Statistical processes have been defined in Section 3.2 (§ 3) as temporal sequences of statistical variables. They may be depicted as

$$Y_0 \longrightarrow Y_1 \longrightarrow Y_2 \longrightarrow Y_3 \longrightarrow \cdots \tag{8.11}$$

These variables have a common range, $\tilde{\mathcal{Y}}$, and are defined alternatively for a fixed reference set, Ω, or for a sequence of changing reference sets, $\Omega_0, \Omega_1, \ldots$ Models of such processes can be descriptive or analytical. Analytical models must be conceived of as population-level models. A stochastic version may be depicted as

$$\ddot{Y}_0^* \longrightarrow\!\!\!\!\!\!\to \dot{Y}_1^* \longrightarrow\!\!\!\!\!\!\to \dot{Y}_2^* \longrightarrow\!\!\!\!\!\!\to \dot{Y}_3^* \longrightarrow\!\!\!\!\!\!\to \cdots$$

It begins with a deterministic population-level variable $\ddot{Y}_0^*$ providing the initial distribution, followed by a sequence of stochastic variables. The arrows represent not just temporal sequence but stochastic functional relationships between population-level variables. As an illustration we briefly discuss diffusion models.

2. Simple diffusion models

Diffusion models concern the spread of some property in a population and start from the idea that the speed of the spread depends in some way on the number of units which have already got the property.[3] A basic version considers a fixed reference set, Ω, and uses statistical variables Y_t such that

$$Y_t(\omega) = \begin{cases} 1 & \text{if } \omega \text{ got the specified property until (and including) } t \\ 0 & \text{otherwise} \end{cases}$$

Consequently, if $Y_t(\omega) = 1$ for the first time, it stays at this value forever. The corresponding individual-level variable will be denoted by $\dot{Y}_t$, and the number of individuals with $\dot{Y}_t = 1$ will be denoted by $\dot{N}_t^*$.[4] The construction of a diffusion model can then start from a multilevel model which, for a single time step, may be depicted as

$$\begin{array}{l} \dot{Y}_t \searrow \\ \qquad\quad \dot{Y}_{t+1} \\ \dot{N}_t^* \nearrow \end{array} \tag{8.12}$$

The state of a generic individual in $t+1$, $\dot{Y}_{t+1}$, stochastically depends on its previous state $\dot{Y}_t$ and on $\dot{N}_t^*$, that is, the current number of individuals who are already in the specified state. A simple specification of the stochastic function is

$$\mathrm{P}(\dot{Y}_{t+1} = 1 \mid \dot{Y}_t = 0, \dot{N}_t^* = n_t) = \alpha \frac{n_t}{n} \tag{8.13}$$

In addition one needs a specification for the generation of $\dot{N}_{t+1}^*$. Assuming that the aggregation results from independent repetitions, one can use

$$\Pr(\dot{N}_{t+1}^* = n_{t+1} \mid \dot{N}_t^* = n_t) = \tag{8.14}$$

$$\binom{n - n_t}{n_{t+1} - n_t} \pi_t^{n_{t+1} - n_t} (1 - \pi_t)^{n - n_{t+1}}$$

with $\pi_t := \alpha\, n_t / n$. The formula provides probabilities for possible developments of the diffusion process. To get an impression of these possible developments, Figure 8.2 shows ten simulated diffusion paths, generated with the algorithm in Box 8.1.

3. Pure individual-level models

The basic idea of a diffusion model is to think of probabilities of individual-level variables as dependent on population-level properties. This requires a multilevel model and an aggregation mechanism as exemplified by (8.13) and (8.14).

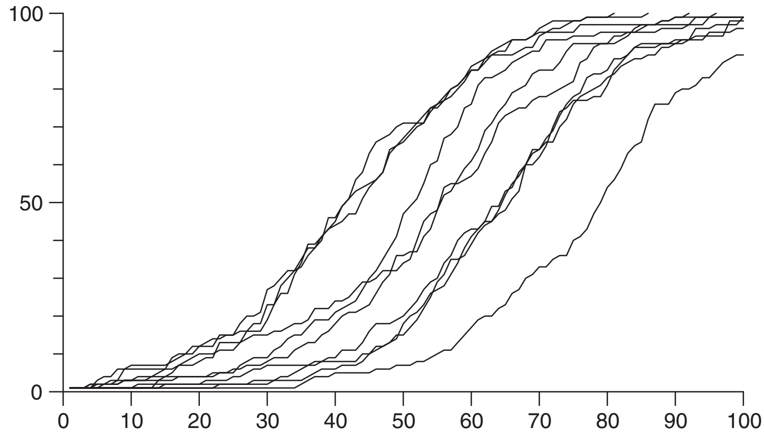


Figure 8.2 Ten simulated diffusion processes, generated with the algorithm shown in Box 8.1. Parameters: $n = 100$, $n_0 = 1$, $\alpha = 0.1$.

Box 8.1 Simulation of diffusion processes according to (8.13) and (8.14).

(1) $t \leftarrow 0$, $n_0 \leftarrow$ initial value
(2) $t \leftarrow t + 1$
(3) $n_t \leftarrow n_{t-1}$; do $(n - n_{t-1})$ times: draw a random number ϵ equally distributed in $[0, 1]$, if $\epsilon \leq \alpha\, n_{t-1}/n$ then $n_t \leftarrow n_t + 1$
(4) if $n_t < n$ continue with (2), otherwise end.

In order to stress the importance of an explicit reference to the population level, I briefly consider the idea to interpret individual-level transition rate models as describing diffusion processes.[5] The approach starts from transition rates $r(t) := \Pr(\dot{Y}_{t+1} = 1 \mid \dot{Y}_t = 0)$. This allows one to derive a survivor function

$$G(t) = \prod_{j=0}^{t-1} (1 - r(j))$$

describing the individual-level process. Estimating such a function from a given set of data, it may also be interpreted as the estimated proportion of individuals not infected until t. On the other hand, if not viewed as a data model, but as a generic functional model, the interpretation of $G(t)$ as a proportion becomes problematic because there is no explicit reference to a population.

One may think of the population, implicitly presupposed by interpreting $G(t)$ as a proportion, as resulting from independent repetitions of the individual-level transition rate model. There remains, however, an important difference to the diffusion model of § 2. Starting from an individual-level transition rate model would allow independent repetitions of individual processes. This would not be possible with a multilevel diffusion model. Even the very simple version described in § 2 would not allow an independent generation of individual processes because time-dependent transition rates depend on outcomes of the population-level process.

4. Modeling interdependencies

It is obvious that diffusion models imply some form of distribution-dependence and, consequently, interdependency among the individual units in the population. It is therefore an interesting question whether these interdependencies can be explicitly represented. One can think in terms of interactions between members of two groups,

$$\mathcal{N}_t^0 := \{\omega \mid Y_t(\omega) = 0\} \quad \text{and} \quad \mathcal{N}_t^1 := \{\omega \mid Y_t(\omega) = 1\}$$

It often seems plausible that a diffusion process depends on contacts between members of these two groups, and moreover on properties of the interacting individuals. However, an explicit representation of such contacts would only be possible if one could refer to identifiable individuals.[6] If this is not possible one can nevertheless follow an idea already hinted at in § 5 of Section 2.2 and use groups (equivalence classes) instead of identifiable individuals.[7]

Assume that $\tilde{\mathcal{X}} = \{\tilde{x}_1, \ldots, \tilde{x}_m\}$ is a property space that allows one to define relations between its categories which can be interpreted as proximities between units in Ω. For example, $\tilde{\mathcal{X}}$ might be a set of spatial locations, and $\delta_{kl} := R(\tilde{x}_k, \tilde{x}_l)$ is some measure of proximity between $\tilde{x}_k$ and $\tilde{x}_l$. Given then a statistical variable $X : \Omega \longrightarrow \tilde{\mathcal{X}}$, one can define groups $\Omega_j := \{\omega \mid X(\omega) = \tilde{x}_j\}$ $(j = 1, \ldots, m)$, and one can assume that δ_{kl} is a measure of proximity between members of Ω_k and Ω_l, respectively.

The population variable $\dot{N}_t^*$ that records the number of units that got the specified property until t can be replaced by a vector $(\dot{N}_{1t}^*, \ldots, \dot{N}_{mt}^*)$, having components $\dot{N}_{jt}^*$ that record the number of units in Ω_j who got the specified property until t. Finally one can generalize (8.13) into

$$\Pr(\dot{Y}_{t+1} = 1 \mid \dot{Y}_t = 0, \dot{X} = \tilde{x}_j, \dot{N}_{1t}^* = n_{1t}, \ldots, \dot{N}_{mt}^* = n_{mt}) = \sum_{k=1}^{m} \delta_{jk}\, \alpha_k \frac{n_{kt}}{n} \tag{8.15}$$

This is now a multilevel diffusion model where the probability of getting the specified property depends on the group, Ω_j, that an individual unit belongs to and, for $k = 1, \ldots, m$, on the value of $\dot{N}_{kt}^*$ and the proximity between Ω_j and Ω_k.

To complete the model, formally analogous to (8.14), one can set up a separate equation for each group:

$$\Pr(\dot{N}^*_{j,t+1} = n_{j,t+1} \,|\, \dot{N}^*_{1t} = n_{1t}, \ldots, \dot{N}^*_{mt} = n_{mt}) = \tag{8.16}$$

$$\binom{n_j - n_{jt}}{n_{j,t+1} - n_{jt}} \pi_{jt}^{n_{j,t+1} - n_{jt}} (1 - \pi_{jt})^{n_j - n_{j,t+1}}$$

with $\pi_{jt} := \sum_k \delta_{jk} \alpha_k n_{kt}/n$ and n_j denoting the size of group j. It is therefore possible to sequentially simulate diffusion processes in basically the same way as was done with the algorithm in Box 8.1. (Illustrations will be discussed in the next section.)

8.3 Functional causality and levels

1. Distribution-dependent causation

Assume a multilevel model as defined in § 5 of Section 8.1 that has the form

$$\begin{array}{l} \ddot{X} \searrow \\ \qquad\quad \dot{Y} \\ \ddot{Z}^* \nearrow \end{array} \tag{8.17}$$

The property spaces $\tilde{\mathcal{X}}$ (of $\ddot{X}$) and $\tilde{\mathcal{Z}}$ (used for $\mathcal{D}[\Omega, \tilde{\mathcal{Z}}]$) may be identical or different. In any case, the functional causal relationship between $\ddot{X}$ and $\dot{Y}$ may depend on values of $\ddot{Z}^*$. The causal relationship will then be called *distribution-dependent*.[8] Obviously, the notion presupposes a multilevel model and cannot be explicated in an individual-level model that does not explicitly refer to a population.

2. Individual effects of changing distributions

Distribution-dependent causation concerns the dependence of a causal relationship, say between $\ddot{X}$ and $\dot{Y}$, on a statistical distribution, say P[Z]. Thinking of possible effects of changes of such distributions, one can distinguish between individual-level and population-level effects. Concerning individual-level effects, the question is how a change in the distribution P[Z] changes the relationship between $\ddot{X}$ and $\dot{Y}$.

The question can be considered in two forms. The example of § 8 in Section 8.1 will be used to illustrate the distinction. One version considers how the conditional probability $\Pr(\dot{Y} = 1 \,|\, \ddot{X} = j, \ddot{X}^* = \mathrm{P}[X])$, defined in (8.10), depends on the distribution of X if $\ddot{X}$ has a fixed value, say $\ddot{X} = j$. An answer can be derived from Figure 8.1 by multiplying the curves with π_j. The solid line in Figure 8.3 shows the result for $s = 50$. The admission probability of a child whose parents have a high educational level gets smaller when the proportion of those children increases.

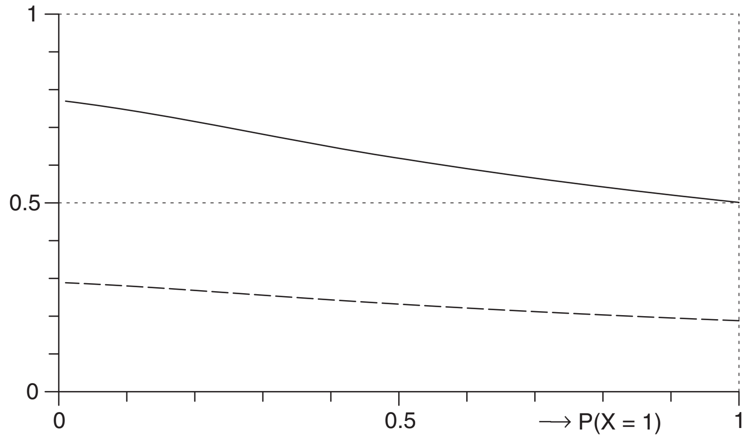


Figure 8.3 Dependence of $\Pr(\dot{Y} = 1 | \ddot{X} = 1, \ddot{X}^* = \mathrm{P}[X])$ (solid line) and $\mathrm{E}(\dot{Y} | \ddot{X} = 1, \ddot{X}^* = \mathrm{P}[X]) - \mathrm{E}(\dot{Y} | \ddot{X} = 0, \ddot{X}^* = \mathrm{P}[X])$ (broken line) on $\mathrm{P}(X = 1)$. Parameters: $n = 100$, $s = 50$, $\pi_0 = 0.5$, and $\pi_1 = 0.8$.

The converse holds for children having parents with a low educational level. If their proportion increases also their admission probability gets larger.

Another version of the question concerns how the effect of a change of $\ddot{X}$ depends on the distribution of X. Accordingly, the broken line in Figure 8.3 shows the dependence of

$$\mathrm{E}(\dot{Y} | \ddot{X} = 1, \ddot{X}^* = \mathrm{P}[X]) - \mathrm{E}(\dot{Y} | \ddot{X} = 0, \ddot{X}^* = \mathrm{P}[X])$$

on $\mathrm{P}(X = 1)$. It is seen that also the comparative advantage of children having parents with a high educational level diminishes if their group extends.

3. Population-level effects

A quite different question concerns effects of changes in statistical distributions at the population level. In the example: How does the proportion of children who get admitted depend on changes in the distribution of parents' educational levels? Since the population-level model is stochastic, there is no unique distribution of the statistical variable Y representing the proportion of admitted children. However, one can calculate a mean proportion as follows:

$$\mathrm{P}(Y = 1) \approx \Pr(\dot{Y} = 1 \,|\, \ddot{X} = 0, \ddot{X}^* = \mathrm{P}[X])\,\mathrm{P}(X = 0) + \Pr(\dot{Y} = 1 \,|\, \ddot{X} = 1, \ddot{X}^* = \mathrm{P}[X])\,\mathrm{P}(X = 1) \tag{8.18}$$

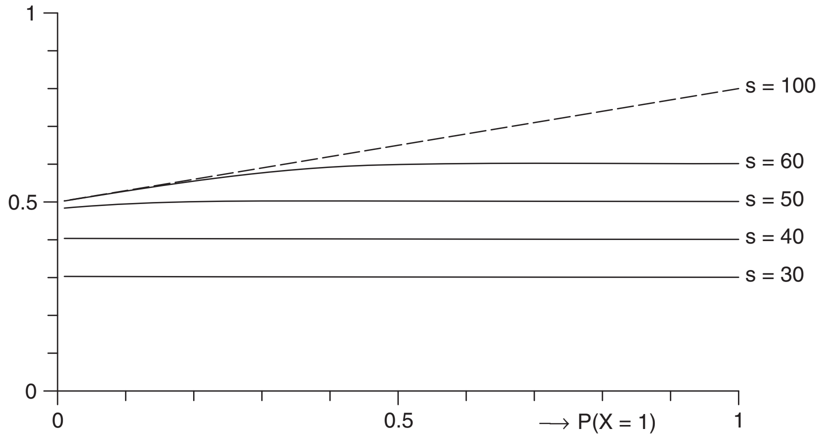


Figure 8.4 Dependence of $P(Y = 1)$ as defined in (8.18) on $P(X = 1)$ for selected values of s. Parameters: $n = 100$, $\pi_0 = 0.5$, and $\pi_1 = 0.8$.

Figure 8.4 shows how this proportion depends on $P(X = 1)$, the proportion of children having parents with a high educational level. It is seen that the relationship is mainly governed by the constraints due to s, the school's capacity. Under a broad variety of circumstances, changes in the distribution of X have no effect for the distribution of Y.

The example not only shows that population-level effects can be quite different from individual-level effects. It also shows that an apparent absence of a relationship at the population level can be the result of counteracting processes at the individual level.

4. *Relationships between levels*

It seems obvious that neither deterministic nor stochastic functional relationships can connect an individual-level variable with a population-level variable.[9] On the other hand, as shown by the previously discussed multilevel models, it is quite possible that a stochastic function connects a population-level variable with an individual-level variable. In fact, both variables may refer to the same property space, e.g. $\dot{Y}_t^* \longrightarrow\!\!\!\!\rightarrow \dot{Y}_{t+1}$ as assumed in a simple diffusion model. The temporal relationship may be left implicit; it is important, however, to be explicit about the process that leads from the population-level to the individual-level variable.[10]

To illustrate, assume that Ω represents children educated in a school. $\dot{Y}$ records a child's educational success (0 or 1), $\ddot{X}$ records the parents' educational level (0 or 1), and the population-level variable $\ddot{X}^*$ provides a statistical distribution, $P[X]$, of the values of $\ddot{X}$ in the school. Assuming that $\dot{Y}$ stochastically depends on $\ddot{X}$ and $\ddot{X}^*$, the model has the structure of (8.5) in Section 8.1. Would it make sense to change $\ddot{X}$ into an endogenous variable $\dot{X}$ resulting in the model

$$\begin{tikzcd} \dot{X} \arrow[dr] & \\ & \dot{Y} \\ \ddot{X}^* \arrow[uu] \arrow[ur] & \end{tikzcd} \tag{8.19}$$

The question is whether, and how, one can think of a process leading from $\ddot{X}^*$ to $\dot{X}$. Of course, knowing a value of $\ddot{X}^*$ would allow to better predict values of $\dot{X}$. But how can one think of a substantial process?

An example would be that the school can select children according to the educational level of their parents. Of course, a selection process cannot change values of $\dot{X}$ for any given child. In fact, the idea of a selection process presupposes a population of children having fixed values of the selection variable. From the point of view of the school, however, the model may well be used to consider causal effects of different selection policies.

5. *Self selection with constraints*

We now reconsider a previously discussed example with self selection (see Section 5.3, § 5): The educational success $\dot{Y}$ (0 or 1) depends on the parents' educational level $\dot{X}$ (0 or 1) and on the school type $\dot{Z}$ (1 or 2), and it is assumed that $\dot{Z}$ is an endogenous variable depending on $\dot{X}$.

How to set up a corresponding multilevel model? One can start from exogenous variables $\ddot{X}^*$ and $\ddot{Z}^*$ providing, respectively, statistical distributions of educational levels and school types. The distribution of school types is taken as exogenous because it does not result from parental choices. Nevertheless, there will be a selection of school types that takes place in the frame of a given distribution $\ddot{Z}^* = \mathrm{P}[Z]$. The selection processes concern the distribution of a variable $(\dot{X}, \dot{Z})$ having marginal distributions given by values of the exogenous population-level variables $\ddot{X}^*$ and $\ddot{Z}^*$. A multilevel model may then be depicted as follows:

$$\begin{tikzcd} \ddot{X}^* \arrow[dr] & & \\ & (\dot{X}, \dot{Z}) \arrow[r] & \dot{Y} \\ \ddot{Z}^* \arrow[ur] & & \end{tikzcd} \tag{8.20}$$

Of course, one needs assumptions about the selection processes that generate the distribution of $(\dot{X}, \dot{Z})$. Here we continue with the numerical illustration of Section 5.3 (§ 5) and assume $\Pr(\dot{Z} = 2 \,|\, \dot{X} = 0) = 0.4$ and $\Pr(\dot{Z} = 2 \,|\, \dot{X} = 1) = 0.8$. However, these are parents' plans, and they might be incompatible with the given distribution of school types. One therefore needs an additional mechanism that solves possible conflicts. For an illustration we simply assume that parents having a high educational level can realize their plans first:

$$\begin{aligned} &\Pr(\dot{X} = 1, \dot{Z} = 2 \,|\, \ddot{X}^* = \mathrm{P}[X], \ddot{Z}^* = \mathrm{P}[Z]) = \\ &\qquad \min\{\Pr(\dot{Z} = 2 \,|\, \dot{X} = 1)\,\mathrm{P}(X = 1), \mathrm{P}(Z = 2)\} \end{aligned} \tag{8.21}$$

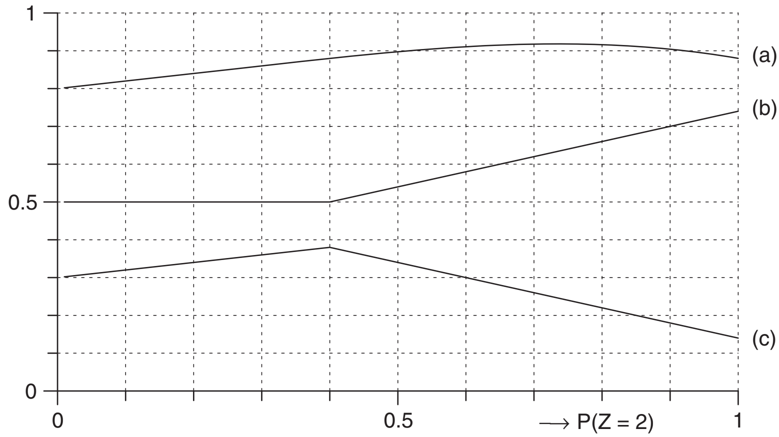


Figure 8.5 Dependence of (a) E_1, (b) E_0, and (c) $E_1 - E_0$, as defined in (8.22), on $P(Z = 2)$ (proportion of schools of type 2).

From this assumption one can calculate the joint distribution of $(\dot{X}, \dot{Z})$ and finally the distribution of $\dot{Y}$.

In particular, one can investigate how the causal effect of a change $\Delta(0, 1)$ in $\dot{X}$, that is, $E_1 - E_0$ with

$$E_j := E(\dot{Y} | \dot{X} = j, \ddot{X}^* = P[X], \ddot{Z}^* = P[Z]) \tag{8.22}$$

depends on the distribution of school types. Based on the numerical specification of the function $(\dot{X}, \dot{Z}) \longrightarrow\!\!\!\!\rightarrow \dot{Y}$ assumed in § 5 of Section 5.3, this is shown in Figure 8.5.

6. Time-dependent effects

Further considerations concern causal relationships in models of statistical processes. An example of the diffusion model introduced in § 4 of Section 8.2 will be used for illustration. In this example there are two groups ($m = 2$), identified by $\tilde{x}_1$ and $\tilde{x}_2$. Parameters are

$$\alpha_1 = 0.2, \quad \alpha_2 = 0.1, \quad \delta_{11} = 1, \quad \delta_{12} = \delta_{21} = 0.1, \quad \delta_{22} = 0.5$$

The population size is $n = 1000$, and the group sizes are $n_1 = 600$ and $n_2 = 400$. Diffusion processes are generated with a slightly modified version of the algorithm in Box 8.1. Figure 8.6 shows one of these processes that starts from $n_{1,0} = n_{2,0} = 1$.

In order to assess effects of group membership one can compare the conditional probabilities (transition rates) defined in (8.15). Figure 8.7 illustrates their development for the example. These are now time-dependent effects. Moreover, the effects heavily depend on the distribution of X. Figure 8.7 was generated with

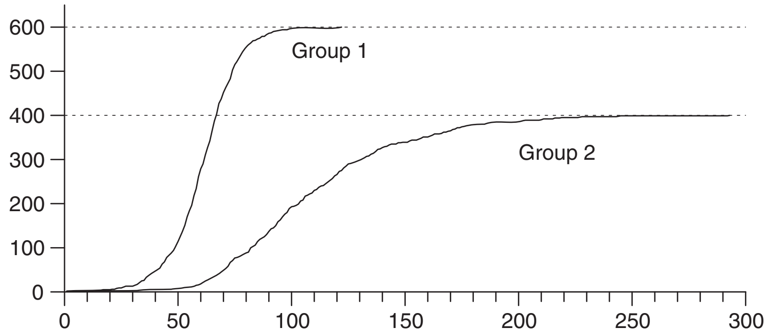


Figure 8.6 Simulation of a structured diffusion processes with two groups. Parameter values are defined at the beginning of § 6. Initial values are $n_{1,0} = n_{2,0} = 1$.

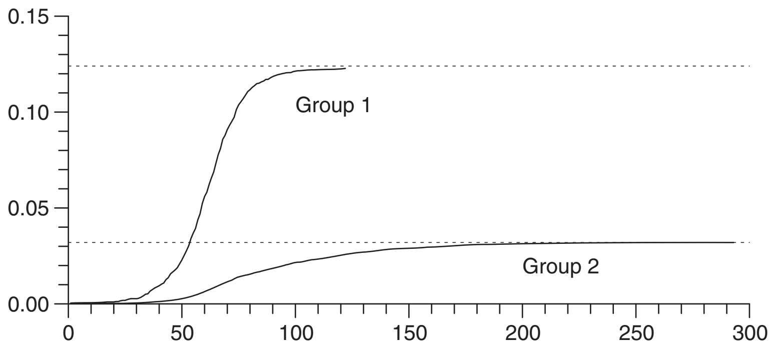


Figure 8.7 Development of the conditional probabilities defined in (8.15), simulated with group sizes $n_1 = 600$ and $n_2 = 400$.

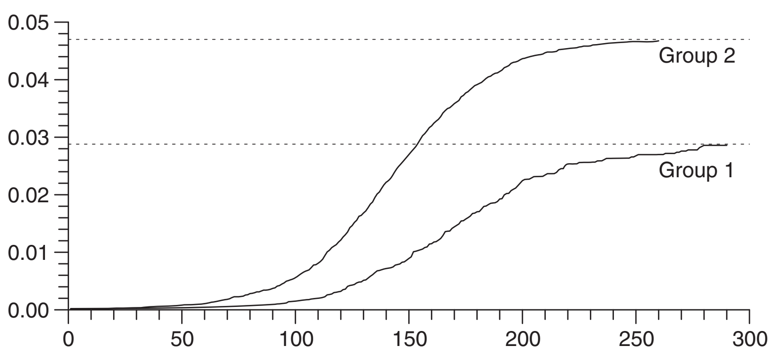


Figure 8.8 Development of the conditional probabilities defined in (8.15), simulated with group sizes $n_1 = 100$ and $n_2 = 900$.

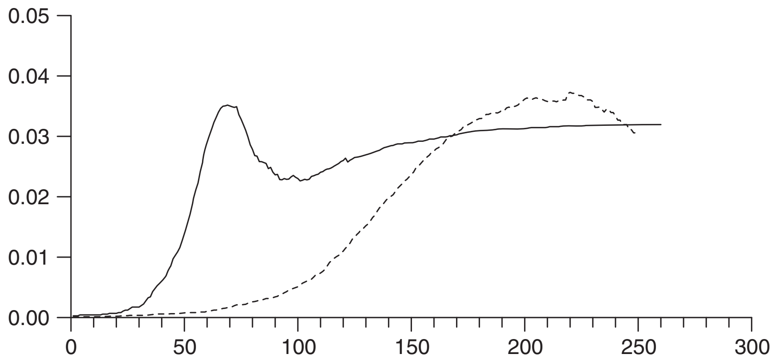


Figure 8.9 Development of the overall rates defined in (8.23). Solid line: $n_1 = 600$ and $n_2 = 400$, dotted line: $n_1 = 100$ and $n_2 = 900$.

group sizes $n_1 = 600$ and $n_2 = 400$. Changing the distribution, e.g. into $n_1 = 100$ and $n_2 = 900$ will result in completely different effects, as shown by Figure 8.8.

7. Mixing group-level effects

The previous paragraph considered the conditional probabilities defined in (8.15) separately for each group. Overall rates, for the whole population, can be derived by mixing the group-specific rates in the following way:

$$\begin{aligned}&\Pr(\dot{Y}_{t+1} = 1 \mid \dot{Y}_t = 0, \dot{N}^*_{1t} = n_{1t}, \ldots, \dot{N}^*_{mt} = n_{mt}) = \\ &\quad \textstyle\sum_j \Pr(\dot{Y}_{t+1} = 1 \mid \dot{Y}_t = 0, \ddot{X} = \tilde{x}_j, \ldots)\,\mathrm{P}(X = \tilde{x}_j \mid \dot{Y}_t = 0)\end{aligned} \tag{8.23}$$

Mixing is with proportions $\mathrm{P}(X = \tilde{x}_j \mid \dot{Y}_t = 0)$, denoting the proportion of members of group $\tilde{x}_j$ in the risk set at t. Figure 8.9 illustrates these overall rates for two different distributions of the group variable X.

Notes

1 Variables and relations

1 In this text a distinction is made between single and double quotation marks. Single quotation marks are used to refer to linguistic expressions; double quotation marks are used either for citations or to indicate that an expression has no clear meaning or that it is used in a metaphorical way. Within citations, it is tried to reproduce quotation marks in their original form. If something is added, or omitted, inside a quotation this will be indicated by square brackets.

2 Throughout this text the word ‘function’ is always used in its mathematical sense.

3 It follows that logical and statistical variables are completely different things. Moreover, the term ‘variable’ is misleading in both cases. For a more extensive discussion that also shows how both notions, logical and statistical variables, can be linked by using sentential functions, see Rohwer and Pötter (2002b: ch. 9). – Of course, when there is no danger of confusion, one can drop the qualification ‘statistical’ and simply speak of variables.

4 In this text, statistical variables will always be denoted by upper case letters (e.g., $A, B, C, \ldots, X, Y, Z$) and their property spaces by corresponding calligraphic letters that are marked by a tilde ($\tilde{\mathcal{A}}, \tilde{\mathcal{B}}, \tilde{\mathcal{C}}, \ldots, \tilde{\mathcal{X}}, \tilde{\mathcal{Y}}, \tilde{\mathcal{Z}}$).

5 In this text ‘$:=$’ is used instead of ‘$=$’ if the definitional character of a statement is to be stressed.

6 If M is a finite set, $|M|$ denotes the number of its elements.

7 One cannot rely on any general rules but needs to consider each statistical calculation in its specific context. As an example, think of household income and rent. Subtracting rent from household income provides a meaningful result, but simply to add both quantities does not.

8 This view has been expressed by the International Statistical Institute (1986: 238) in a *Declaration on Professional Ethics* in the following way: “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.”

9 In this text, $\mathbf{R}$ is used to denote the set of real numbers, and $\mathcal{P}(\tilde{\mathcal{X}})$ denotes the power set of $\tilde{\mathcal{X}}$, that is, the set of all subsets of $\tilde{\mathcal{X}}$.

10 Of course, different notations can be used inside the round brackets to indicate the subsets of the variable’s property space to be used as arguments.

11 The two notations will be used equivalently. While the notation on the right-hand side of the first equality sign is formally preferable because it clearly distinguishes between the name of the function and possibly appended arguments, the notation used on the left-hand side often allows an easier grasp of the intended meaning.

12 Arguments behind the semicolon will be treated as parameters of the function.

13 In order to keep the illustration simple household size will not be included as a separate variable.
14 See, e.g., Burt (1982: 4–8), Wellman (1988: 38–9).
15 Explicitly written: $\Omega \times \Omega = \{(\omega, \omega') \mid \omega, \omega' \in \Omega\}$.
16 "The events are various: a day's work behind the counter of a store, a meeting of a women's club, a church supper, a card party, a supper party, a meeting of the Parent–Teacher Association, etc." (Homans 1951: 82).
17 For an extensive overview see Wasserman and Faust (1994).
18 Symmetry can be defined in an obvious way for arbitrary relational variables. The other two properties are only defined if $\tilde{\mathcal{R}} = \{0, 1\}$: R is reflexive if $R(\omega, \omega) = 1$ for all $\omega \in \Omega$; and R is transitive if, for all $\omega, \omega', \omega'' \in \Omega$, $R(\omega, \omega') = R(\omega', \omega'') = 1$ implies $R(\omega, \omega'') = 1$.

2 Notions of structure

1 For a detailed critique see Bates and Peacock (1989).
2 Note that the term 'data generating process' is sometimes also used to refer to substantial processes. The following statement made by D. Freedman (1985: 348) provides an example: "In social-science regression analysis [...] usually the idea is to fit a curve to the data, rather than figuring out the process which generated the data. As a matter of fact, investigators often talk about 'modeling the data.' This is almost perverse: surely the object is to model the phenomenon, and the data are interesting only because they contain information about that phenomenon." Freedman obviously means that one should be interested in the substantial processes that created the facts (referred to by the data).
3 This is not to deny that institutions exist (e.g., voting systems) which employ methods for an aggregation of individual behavior that are formally similar to statistical methods. The aggregated outcome can then sensibly be characterized as resulting from the behavior of individual actors, but is clearly not just a statistical fact.
4 This should not be confused with the assumption, often made when estimating stochastic models, that the individual observations contained in a sample can be considered as corresponding to independent repetitions of the model. See, e.g., Greene (1993: 87).
5 See also the discussion, in Przeworski and Teune (1970), of using proper names in comparative research.
6 See Wasserman and Faust (1994: 357).
7 A permutation $\pi : \Omega \longrightarrow \Omega$ is called an automorphism *iff* $R(\omega', \omega'') = R(\pi(\omega'), \pi(\omega''))$ for all $\omega', \omega'' \in \Omega$. See Everett *et al.* (1990).
8 Note that this statement does not invalidate the remark, made in § 1 of Section 2.1, that the statistical notion of structure has no relational connotations. The additional argument only says that it might well be possible to represent (aspects of) a relational structure by statistical variables; and this then would allow one to make statistical statements about the relational structure (e.g., about the distribution of living alone or involved in a relationship).
9 See, e.g., Knoke and Kuklinski (1982: 9), Berkowitz (1988: 480–1), Wasserman and Faust (1994: 4).
10 Obviously, if ω' and ω'' are statistically equivalent w.r.t. X, they are structurally equivalent w.r.t. R^*.

3 Processes and process frames

1 For a good discussion see the essay "Present, Future and Past" in Oakeshott (1983).
2 A similar distinction has been proposed by 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 ein subjektives Schema narrativer Aneignung, sondern dessen äußere Bedingung."

3 See the discussion of this distinction in B. C. Smith (1996: 32–6).

4 In this text, a process is termed *mechanical* if human actors are not involved or can only initialize the process without influencing its subsequent development. Using this understanding the statement made above does not include interactive computer programs.

5 Although in many applications all logical variables have the same range, this is not required by the definition.

6 Models which consider X_t^* as modal variables will be discussed in Chapter 8.

7 If such processes are considered as repeatable one also needs suitably defined process frames; this will be discussed when dealing with population-level models in Chapter 8.

8 The explanation here given follows the treatment in Rohwer and Pötter (2002b).

9 These are therefore prototypical examples of mechanical processes as defined in Section 3.1 (§ 3, footnote 4).

4 Functional models

1 Moreover, in order to think of Ω as a random sample from $\Omega^{\dagger}$, all elements of the population must exist at the time of generating the sample.

2 The designation reflects the very broad usage of the term 'causal' in the scientific literature. It is certainly reasonable, and will be done in Chapter 5 to distinguish different understandings and definitions of causality.

3 As explained in Section 3.3, in order to distinguish *deterministic modal variables* from stochastic modal variables as well as from statistical variables they are designated by two points.

4 If there is no danger of confusion, qualifications of variables will be dropped.

5 As for statistical variables, it will be assumed for modal variables that there exists a numerical representation for their values; ranges of modal variables will therefore be treated as subsets of the set of real numbers.

6 This is the distinct feature of deterministic models; stochastic versions will be introduced in Section 4.2.

7 The meaning of 'effective condition' will be explained below in § 9.

8 This is easily seen if one views $\mathrm{B}_{\mathcal{M}}[\ddot{X}] \times \mathrm{B}_{\mathcal{M}}[\ddot{Y}]$ as a two-dimensional table. It is assumed that both domains, $\mathrm{B}_{\mathcal{M}}[\ddot{X}]$ as well as $\mathrm{B}_{\mathcal{M}}[\ddot{Y}]$, contain at least two elements.

9 As an example, consider the following situation where $\ddot{X}$ and $\ddot{X}'$ represent two parallel switches for the bulb $\ddot{Y}$:

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

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

10 Consider, for example, the following situation:

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

$\ddot{X}$ and $\ddot{Y}$ are independent on the left-hand side, but dependent on the right-hand side.

11 In this text, talk of constraints is always meant to imply that the resulting dependencies cannot be expressed by functions involving only the given variables.
12 It suffices to specify a distinct value of $\ddot{Z}$ for each pair $(x, y) \in \mathrm{B}_{\mathcal{M}}[\ddot{X}, \ddot{Y}]$.
13 Deterministic relationships of the form $\ddot{X} \longrightarrow \dot{Y}$ or $\dot{X} \longrightarrow \ddot{Y}$ are not possible.
14 If the function is called f: $\Pr(\dot{Y}=y) = \sum_{x \in f^{-1}(\{y\})} \Pr(\dot{X}=x)$.
15 Whether deterministic or stochastic variables are used to formulate the conditions is obviously irrelevant. Of course, only with a stochastic variable would it be possible to form a common distribution with the dependent variable. It is important, however, that the stochastic function $x \longrightarrow \Pr[\dot{Y} | \dot{X}=x]$ is completely independent of any assumptions about the distribution of $\dot{X}$ and, in particular, does not require that all values in the range of $\dot{X}$ can occur with a positive probability.
16 Since $\ddot{X}$ is a deterministic variable one might prefer to say that $\dot{Y}$ is stochastically dependent on, or independent of, $\ddot{X}$.
17 Obviously, one can also define an asymmetric notion of functional dependence as was done in Section 4.1 (§ 9).
18 A stochastic variable which is stochastically dependent on an exogenous variable is then, by definition, an endogenous variable; see also Section 4.3 (§ 6).
19 See, e.g., Spirtes *et al.* (1993), Cooper (1999), Greenland *et al.* (1999), Robins (1999), Pearl (2000), Woodward (2001: 41).
20 For simplicity, this formulation presupposes that the model does not contain ultimate deterministic variables.
21 Beginning with a model of the form (4.4), the construction of (4.5) can always be done in a trivial way: Using as the range of $\dot{Z}$ the Cartesian product of the ranges of $\ddot{X}$ and $\dot{Y}$, the distribution of $\dot{Z}$ can be defined by the conditional probabilities $\Pr(\dot{Y}=y \,|\, \ddot{X}=x)$.
22 In general, pseudo-indeterministic models are models containing only deterministic functions and at least one stochastic endogenous variable. Their stochastic features are completely determined from exogenous stochastic variables. See, e.g., Glymour *et al.* (1991: 155–6), Spirtes *et al.* (1993: 38–9), and Papineau (2001: 17).
23 See, e.g., Pearl (2000: 26–7, 44), and Woodward (2003: 339).
24 Such questions are rarely discussed in the literature; but see Tryfos (2004: ch. 5).
25 See, e.g., Freedman (1992), Clogg and Haritou (1997), Woodward (2003: 325–7).
26 The situation is quite different for defined unobserved variables. Questions concerning possible dependencies may then become quite important, in particular for causal interpretations. This will be discussed in Chapter 5.

5 Functional causality

1 The notation $\Delta(x', x'')$ is used to express that the value of a variable, referred to in the context, changes from x' to x''. The notation implies that $x' \neq x''$.
2 Instead of $\ddot{Z}$ one can also use a vector $(\ddot{Z}_1, \ldots, \ddot{Z}_m)$ who's components are elements of $\mathcal{V}$.
3 This is in accordance with the idea that causal statements are (most often) intended to capture relationships between events, or types of events; see, e.g., Hausman (1998: ch. 2).
4 This text therefore avoids an unspecific talk of "(causal) factors," in particular, in contexts where causes and (functional) conditions might be confused. See also Woodward (2003: 39).
5 Beginning with John St. Mill, this has led to a quite different notion of causality which identifies causes and conditions, see, e.g., Rothman and Greenland (2005), and Susser (1991). This notion should not be confused with the definition given above.
6 For another example that clearly illustrates the distinction see Hausman (1998: 25).
7 See also the remarks in § 6. A dynamic version of functional causality that conceives of causes as events will be discussed in Section 7.3.

8 See, e.g., Marini and Singer (1988: 354). The main promoter of this idea is, again, John St. Mill who explicitly identified causes and conditions.

9 This immediately avoids several problems that easily occur if one thinks instead, in some obscure sense, of "factors;" see the discussion in Hitchcock (1993).

10 Such attempts were widespread in the early discussions of probabilistic causality, see, e.g., Eells and Sober (1983: 37), and Eells (1991: 94–107). Against these attempts, the importance of context-dependencies has been stressed by Carroll (1991). For continual discussion of the "contextual unanimity thesis" see Twardy and Korb (2004).

11 This was assumed by Suppes (1970) in his original definition of probabilistic causes and then adopted, e.g., by Sober (1986: 97). For further comments on this assumption see Eells (1988, 1991).

12 This idea already played an important role in Suppes' (1970) original discussion of probabilistic causality. The same idea also shows up in a widespread statistical notion of causality as can be seen from the following formulation given by Cox (1992: 293): "[...] 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 also provides a critical discussion; for further discussion from a sociological point of view see Goldthorpe (2001).

13 See, e.g., Holland (1986: 959; 1988b; 1993).

14 Rubin (1974), Rosenbaum and Rubin (1983), Holland (1986, 1988a, 1994, 2001). Adaptations for social science applications have been propagated by King *et al.* (1994), Sobel (1995), and Winship and Morgan (1999).

15 The discussion will be continued in Section 6.2.

16 See Dawid (2000), the discussion which followed his contribution, and also the additional comments in Dawid (2006).

17 These models have been called "causal models" by Holland (2001: 178).

18 Characterizing this as a modal question means that the question concerns possibilities. Counterfactual questions are modal questions of a special type, based on counterfactual presuppositions; for example: How might the disease have developed if, contrary to what actually was done, the therapy would not have been applied? For a discussion of some of these variants of modal questions see Dawid (2000, 2006).

19 A stochastic version of this model that uses stochastic variables instead of $\ddot{X}$ and $\ddot{I}$ was proposed by Steyer *et al.* (2000).

20 This implies that functions depending non-trivially on the identity of the individuals are ruled out.

21 See Holland (1986: 948).

22 The range is specified by the set of real numbers, **R**, because expected values of $\dot{Y}$ need not be elements of $\tilde{\mathcal{Y}}$.

23 See, e.g., McKim (1997: 7–8), Wooldridge (2002: 3), or Bollen (1989: 41).

24 Heckman (2005: 1) wrote: "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." See also his definition of functional causality in Heckman (2000: 52–3).

25 Note that the definition of (5.9) nevertheless requires that $\dot{Z}$ is stochastically independent of $\ddot{X}$.

26 In both models, (5.7) and (5.11), it is not possible to hold constant the values of $\ddot{X}$ if a change $\Delta(z', z'')$ occurs. However, in model (5.11) it is not necessary to consider $\ddot{X}$ as a covariate context for effects of $\Delta(z', z'')$.

27 One can use the equation

$$\begin{aligned}&\mathrm{E}(\dot{Y}|\ddot{X}=x) = \\ &\quad \mathrm{E}(\dot{Y}|\ddot{X}=x,\dot{Z}=1)\,\Pr(\dot{Z}=1|\ddot{X}=x) + \mathrm{E}(\dot{Y}|\ddot{X}=x,\dot{Z}=2)\,\Pr(\dot{Z}=2|\ddot{X}=x)\end{aligned}$$

to perform the calculations.

6 Models and statistical data

1 In the literature these models are often called "statistical models." There is no uniform usage of this expression. The naming conventions followed here correspond to the distinction between statistical and stochastic variables.
2 In the statistical literature, authors using stochastic data models often follow this idea. The following quote from Cox and Wermuth (1996: 12) provides an example: "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 mathematical 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." What the authors call a probability model obviously corresponds to a stochastic data model as explained above.
3 That is, $\mathrm{P}[Y|X=x, Z=z]$ serves to estimate $\Pr[\dot{Y}|\ddot{X}=x, \ddot{Z}=z]$.
4 It is noteworthy that this fact becomes obscured if one begins with a closed functional model that presupposes specified probability distributions for all exogenous variables. As an example, see the modeling approach in Woodward (2001: 41).
5 In a formulation of Holland (1994: 269): "Randomization is a physical act in which a known chance mechanism is used in particular ways to construct the function x [that is, in our notation, the statistical variable X]." For an extensive discussion of different methods see Shadish *et al.* (2002).
6 A possible argument would be: If values of (X, Z) could be observed then, as a consequence of the randomization, X and Z would be approximately statistically independent.
7 In order to conform to the notations of the present section the meaning of $\ddot{X}$ and $\ddot{Z}$ has been reversed.
8 See also the critical remarks made by Cox and Wermuth (2004: 292).
9 Note that it must be assumed, in order to set up the table, that values of the variables Y_x exist independently of the actually applied treatment.
10 This proposal was already made by Neyman who is often viewed as one of the originators of the potential outcome idea; see Splawa-Neyman (1990) and the comment by Rubin (1990).
11 Note that the problem does not result from Holland's deterministic approach that is followed here. Essentially the same problem, and the same inconsistencies, would result if the fictitious variables, Y_x and $Y'_{x'}$, would be conceived of as stochastic quantities.
12 The argument does not exclude the possibility to combine the causes. In the example one could combine X and X' into a variable X'' and for each of its four possible values assume a specified educational success. It would then not be possible, however, to associate with each of its components, X and X', separate causal effects; and, in particular, separate randomizations with respect to the components would not be possible.
13 See, e.g., Collier *et al.* (2004: 230–3).
14 This idea is followed, for example, by Rosenbaum (2002).
15 Conceptualizations of processes by which causes come into being depend on whether one refers to dynamic or comparative causes; but for the current consideration it suffices to abstract from the distinction.
16 For a similar argument see Heckman (1992).
17 For systematic expositions of this approach see, in particular, Woodward (2003), and Pearl (2000).

18 Woodward (2001: 50); Pearl (1998: 266, 1999: 105–6), Scheines (1997: 189).
19 See Woodward (2003: 98), Pearl (2000: 70). Somewhat different formulations have been discussed by Dawid (2002).
20 In many cases the modified model then shows that effects of the supposed cause are context-dependent. This problem is rarely discussed because most authors confine their considerations to closed models. Starting with a model only containing stochastic variables, an intervention with respect to a variable $\dot{X}$ creates a modified model with only a single exogenous variable $\ddot{X}$ instead of $\dot{X}$. In this way it may seem possible to think of uniquely defined mean effects due to a change $\Delta(x', x'')$.
21 See Nurmi (1974).
22 See also Goldthorpe (2001: 8).
23 See Section 6.2, § 4.
24 This belief is often already by the definition a part of the notion of a hypothetical experiment, see, e.g., Woodward (2003: 96).
25 This has been stressed by Heckman (1996: 461).
26 See, e.g., Sobel (2005: 117).
27 See, e.g., Maddala (1977, 1983), Heckman (1990), Wooldridge (2002).
28 This is the topic of (probabilistic) selection models; for a survey see Pötter (2006).
29 Just in order to follow Maddala's argument one might accept the unreal presuppositions that people can deliberately decide for or against a college education and that their decisions only depend on expected income.
30 See Maddala (1977: 354).
31 Of course, one also needs the assumption that the distributions of $\dot{Y}_0$ and $\dot{Y}_1$ are in some sense given and known because they cannot be estimated from observations of $\dot{D}$ and $\dot{Y}$.
32 Another possibility would be to substitute the decision rule (6.21) by a randomization device. The modified model then contains an exogenous stochastic variable $\dot{D}$ (having a known distribution), and there is no longer an arrow from $(\dot{Y}_0, \dot{Y}_1)$ to $\dot{D}$.
33 See Hausman (1998: ch. 11), Woodward (1999).

7 Models with event variables

1 These notations specify event variables as stochastic variables. Analogously, the notation $\ddot{E}, \ddot{E}_1, \ddot{E}_2, \ldots$ is used to refer to deterministic event variables.
2 In the following, the formulation that an event variable *takes a specific value* always means that the variable takes a value that is not equal to zero associated with a specific event type.
3 The | and the ‖ notation for conditional probabilities differentiates between distinct meanings; the formal properties are the same.
4 Without qualification contextual variables are assumed to be static. Instead of an exogenous state variable $\ddot{X}$ one could use a deterministic event variable. The model then refers to a situation where this event variable already has a specific value.
5 The variants correspond to (6.14) in Section 6.2.
6 Of course, one can imagine a situation in which the event variable $\ddot{E}_1$ gets a specific value. However, because $\ddot{E}_1$ is exogenous the situation would not be a part of the model.
7 For an empirical investigation of such situations see Blossfeld *et al.* (1999).
8 Another, and often simpler, approach would employ binary event variables, for example $(\dot{E}_1, \dot{E}_2, \dot{E}_3)$, with $\dot{E}_1 = 1$ for a marriage, $\dot{E}_2 = 1$ for a pregnancy, and $\dot{E}_3 = 1$ for a separation. There is then a first situation where one of the events

$$(1, 0, 0), (1, 1, 0), (0, 1, 0), (0, 1, 1), (0, 0, 1)$$

can occur and in the first three cases one can consider successive situations.

9 A version of the model without the contextual variable $\ddot{Z}$ can be illustrated by the following numerical values:

$$\begin{array}{ll}
\sigma_1 : \Pr(\dot{E}_1 = 1, \dot{E}_2 = 0) = 0.30 & \sigma_{21} : \Pr(\dot{E}_1 = 1 \,\|\, \dot{E}_2 = 1) = 0.6 \\
\sigma_1 : \Pr(\dot{E}_1 = 2, \dot{E}_2 = 0) = 0.48 & \sigma_{21} : \Pr(\dot{E}_1 = 2 \,\|\, \dot{E}_2 = 1) = 0.4 \\
\sigma_1 : \Pr(\dot{E}_1 = 0, \dot{E}_2 = 1) = 0.15 & \sigma_{22} : \Pr(\dot{E}_2 = 1, \dot{E}_3 = 0 \,\|\, \dot{E}_1 = 1) = 0.69 \\
\sigma_1 : \Pr(\dot{E}_1 = 1, \dot{E}_2 = 1) = 0.03 & \sigma_{22} : \Pr(\dot{E}_2 = 0, \dot{E}_3 = 1 \,\|\, \dot{E}_1 = 1) = 0.28 \\
\sigma_1 : \Pr(\dot{E}_1 = 2, \dot{E}_2 = 1) = 0.03 & \sigma_{22} : \Pr(\dot{E}_2 = 1, \dot{E}_3 = 1 \,\|\, \dot{E}_1 = 1) = 0.03
\end{array}$$

10 It follows that the expression $\tau(\dot{E})$ has a defined value only if there is already a specific value for the event variable used as an argument.

11 There are other approaches to dynamic causality that conceive of events as temporally identifiable changes of the values of state variables. For a discussion of these approaches see Pötter and Blossfeld (2001).

12 Of course, there can be cases that do not require a reference to any specific context.

8 Multilevel and population-level models

1 See the critical remarks made by Coleman (1990: 1).

2 Explicitly indicating Ω in the notation emphasizes its importance in the definition. It implies, e.g., that the number of units is known.

3 See, e.g., Bartholomew (1982), Mahajan and Peterson (1985), Rogers (1995), Morris (1994), Palloni (2001).

4 Notice that $\dot{N}_t^*$ is derived, not from $\dot{Y}_t$, but from the population-level variable $\dot{Y}_t^*$ recording the distribution of Y_t in Ω.

5 See, e.g., Diekmann (1989), Brüderl and Diekmann (1995).

6 This is assumed in the approaches proposed by Strang (1991), Strang and Tuma (1993), Greve *et al.* (2001), Yamaguchi (1994), Buskens and Yamaguchi (1999). These models therefore relate to diffusion processes in structured units.

7 To be sure, the idea is not new but has a long history in modeling structured diffusion processes; see Morris (1994).

8 The special case where $\ddot{Z}^*$ refers to a statistical distribution of values of $\ddot{X}$ is often called *frequency-dependent causation*, following Sober (1982).

9 This does not exclude the possibility that selecting the value of an individual-level variable may constrain the range of possible values of a (corresponding) population-level variable.

10 This has been stressed in the literature dealing with modeling contextual effects; see, e.g., Blalock (1984), Duncan and Raudenbush (1999).

References

Barker, S. (2004) "Analysing Chancy Causation without Appeal to Chance-raising", *Cause and Chance*, ed. P. Dowe and P. Noordhof, London: Routledge, 120–37.

Bartholomew, D. J. (1982) *Stochastic Models for Social Processes,* 3rd edn, New York: Wiley.

Bates, F. L. and Peacock, W. G. (1989) "Conceptualizing Social Structure: The Misuse of Classification in Structural Modeling", *American Sociological Review* 54, 565–77.

Berkowitz, S. D. (1988) "Afterword: Toward a Formal Structural Sociology", *Social Structures: A Network Approach*, ed. B. Wellman and S. D. Berkowitz, Cambridge: Cambridge University Press, 477–97.

Blalock, H. M. (1984) "Contextual-Effects Models: Theoretical and Methodological Issues", *Annual Review of Sociology* 10, 353–72.

Blau, P. M. (1974) "Parameters of Social Structure", *Approaches to the Study of Social Structure* (1976), London: Open Books, 220–53.

Blau, P. M. (1994) *Structural Contexts of Opportunities*, Chicago: University of Chicago Press.

Blossfeld, H.-P., Klijzing, E., Pohl, K. and Rohwer, G. (1999) "Why Do Cohabiting Couples Marry? An Example of a Causal Event History Approach to Interdependent Systems", *Quality & Quantity* 33, 229–42.

Bollen, K. A. (1989) *Structural Equations with Latent Variables*, New York: Wiley.

Brüderl, J. and Diekmann, A. (1995) "The Log-Logistic Rate Model", *Sociological Methods & Research* 24, 158–86.

Burt, R. S. (1982) *Toward a Structural Theory of Action*, New York: Academic Press.

Buskens, V. and Yamaguchi, K. (1999) "A New Model for Information Diffusion in Heterogenous Social Networks", *Sociological Methodology* 29, 281–325.

Carroll, J. W. (1991) "Property-Level Causation?", *Philosophical Studies* 63, 245–70.

Cartwright, N. (1979) "Causal Laws and Effective Strategies", *Noûs* 13, 419–37.

Clogg, C. C. and Haritou, A. (1997) "The Regression Method of Causal Inference and a Dilemma Confronting This Method", *Causality in Crisis? Statistical Methods and the Search for Causal Knowledge in the Social Sciences*, ed. V. R. McKim and S. P. Turner, Notre Dame: University of Notre Dame Press, 83–112.

Coleman, J. S. (1990) *Foundations of Social Theory*, Cambridge: Harvard University Press.

Collier, D., Brady, H. E. and Seawright, J. (2004) "Sources of Leverage in Causal Inference: Toward an Alternative View of Methodology", *Rethinking Social Inquiry*, ed. H. E. Brady and D. Collier, New York: Rowman & Littlefield, 229–66.

Cooper, G. F. (1999) "An Overview of the Representation and Discovery of Causal Relationships Using Bayesian Networks", *Computation, Causation, and Discovery*, ed. C. Glymour and G. F. Cooper, Cambridge: MIT Press, 3–62.

Cox, D. R. (1992) "Causality: Some Statistical Aspects", *Journal of the Royal Statistical Society* A 155, 291–301.

Cox, D. R. and Wermuth, N. (1996) *Multivariate Dependencies – Models, Analysis and Interpretation*, London: Chapman & Hall.

Cox, D. R. and Wermuth, N. (2004) "Causality: a Statistical View", *International Statistical Review* 72, 285–305.

Dawid, A. P. (2000) "Causal Inference without Counterfactuals", *Journal of the American Statistical Association* 95, 407–24.

Dawid, A. P. (2002) "Influence Diagrams for Causal Modelling and Inference", *International Statistical Review* 70, 161–89.

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.

Diekmann, A. (1989) "Diffusion and Survival Models for the Process of Entry to Marriage", *Journal of Mathematical Sociology* 14, 31–44.

Duncan, G. J. and Raudenbush, S. W. (1999) "Assessing the Effects of Context in Studies of Child and Youth Development", *Educational Psychologist* 34, 29–41.

Eells, E. (1988) "Probabilistic Causal Interaction and Disjunctive Causal Factors", *Probability and Causality*, ed. J. H. Fetzer, Dordrecht: Reidel, 189–209.

Eells, E. (1991) *Probabilistic Causality*, Cambridge: Cambridge University Press.

Eells, E. and Sober, E. (1983) "Probabilistic Causality and the Question of Transitivity", *Philosophy of Science* 50, 35–57.

Elwert, F. and Winship, C. (2002) "Commentary: Population versus Individual Level Causal Effects", *International Journal of Epidemiology* 31, 432–4.

Everett, M. G., Boyd, J. P. and Borgatti, S. P. (1990) "Ego-Centered and Local Roles: A Graph Theoretic Approach", *Journal of Mathematical Sociology* 15, 163–72.

Fisher, R. A. (1922) "On the Mathematical Foundations of Theoretical Statistics", *Philosophical Transactions of the Royal Society of London*, Series A, Vol. 222, 309–68.

Freedman, D. A. (1985) "Statistics and the Scientific Method", *Cohort Analysis in Social Research*, ed. W. M. Mason and S. E. Fienberg, New York: Springer-Verlag, 343–66.

Freedman, D. A. (1992) "As Others See Us: A Case Study in Path Analysis", *The Role of Models in Nonexperimental Social Science: Two Debates*, ed. J. P. Shaffer, Washington: American Educational Research Association, 3–30.

Glenn, N. D. (1977) *Cohort Analysis*, Beverly Hills: Sage.

Glymour, C., Spirtes, P. and Scheines, R. (1991) "Causal Inference", *Erkenntnis* 35, 151–89.

Goldthorpe, J. H. (2001) "Causation, Statistics, and Sociology", *European Sociological Review* 17, 1–20.

Greene, W. H. (1993) *Econometric Analysis*, 2nd edn, New York: Macmillan.

Greenland, S., Pearl, J. and Robins, J. M. (1999) "Causal Diagrams for Epidemiologic Research", *Epidemiology* 10, 37–48.

Greve, H. R., Tuma, N. B. and Strang, D. (2001) "Estimating Diffusion Processes from Incomplete Data", *Sociological Methods & Research* 29, 435–67.

Hausman, D. M. (1998) *Causal Asymmetries*, Cambridge: Cambridge University Press.

Heckman, J. J. (1990) "Selection Bias and Self-selection", *The New Palgrave: Econometrics*, ed. J. Eatwell, M. Milgate and P. Newman, New York: W. W. Norton & Comp., 201–24.

Heckman, J. J. (1992) “Randomization and Social Policy Evaluation”, *Evaluating Welfare and Training Programs*, ed. C. F. Manski and I. Garfinkel, Cambridge: Harvard University Press, 201–30.

Heckman, J. J. (1996) “Comment”, *Journal of the American Statistical Association* 91, 459–62.

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.

Hernes, G. (1976) “Structural Change in Social Processes”, *American Journal of Sociology* 82, 513–47.

Hitchcock, C. R. (1993) “A Generalized Probabilistic Theory of Causal Relevance”, *Synthese* 97, 335–64.

Holland, P. W. (1986) “Statistics and Causal Inference”, *Journal of the American Statistical Association* 81, 945–68.

Holland, P. W. (1988a) “Causal Inference, Path Analysis, and Recursive Structural Equation Models”, *Sociological Methodology 1988*, ed. C C. Clogg, San Francisco: Jossey-Bass, 449–84.

Holland, P. W. (1988b) “Comment: Causal Mechanism or Causal Effect: Which is Best for Statistical Science?”, *Statistical Science* 3, 186–89.

Holland, P. W. (1993) “Which Comes First, Cause or Effect?”, *A Handbook for Data Analysis in the Behavioral Sciences*, ed. G. Keren and C. Lewis, Hillsdale: Lawrence Erlbaum, 273–82.

Holland, P. W. (1994) “Probabilistic Causality Without Probability”, *Patrick Suppes: Scientific Philosopher, Vol. 1*, ed. P. Humphreys, Dordrecht: Kluwer, 257–92.

Holland, P. W. (2001) “The Causal Interpretation of Regression Coefficients”, *Stochastic Causality*, ed. M. C. Galavotti, P. Suppes and D. Costantini, Stanford: Center for the Study of Language and Information, 173–87.

Homans, G. C. (1951) *The Human Group*, London: Routledge & Kegan Paul.

Homans, G. C. (1976) “What Do We Mean by Social “Structure”?”, *Approaches to the Study of Social Structure*, ed. P. M. Blau, London: Open Books, 53–65.

International Statistical Institute (1986) “Declaration on Professional Ethics”, *International Statistical Review* 54, 227–42.

Jauss, H. (1973) “Versuch einer Ehrenrettung des Ereignisbegriffs”, *Geschichte – Ereignis und Erzählung*, ed. R. Koselleck and W. D. Stempel, München: Fink, 554–60.

King, G., Keohane, R. O. and Verba, S. (1994) *Designing Social Inquiry. Scientific Inference in Qualitative Research*, Princeton: Princeton University Press.

Knoke, D. and Kuklinski, J. H. (1982) *Network Analysis*, Newbury Park: Sage.

McKim, V. R. (1997) “Introduction”, *Causality in Crisis? Statistical Methods and the Search for Causal Knowledge in the Social Sciences*, ed. V. R. McKim and S. P. Turner, Notre Dame: University of Notre Dame Press, 1–19.

Maddala, G. S. (1977) “Self-Selectivity Problems in Econometric Models”, *Applications of Statistics*, ed. P. R. Krishnaiah, Amsterdam: North-Holland, 351–66.

Maddala, G. S. (1983) *Limited-dependent and Qualitative Variables in Econometrics*, Cambridge: Cambridge University Press.

Mahajan, V. and Peterson, R. A. (1985) *Models for Innovation Diffusion*, Beverly Hills: Sage.

Marini, M. M. and Singer, B. (1988) "Causality in the Social Sciences", *Sociological Methodology 1988*, ed. C. C. Clogg, San Francisco: Jossey-Bass, 347–409.
Morris, M. (1994) "Epidemiology and Social Networks: Modeling Structured Diffusion", *Advances in Social Network Analysis*, ed. S. Wasserman and J. Galaskiewicz, Thousand Oakes: Sage, 26–52.
Nurmi, H. (1974) "Social Causality and Empirical Data Reduction Techniques", *Quality and Quantity* 8, 159–80.
Oakeshott, M. (1983) *On History and Other Essays*, Oxford: Basil Blackwell.
Palloni, A. (2001) "Diffusion in Sociological Analysis", *Diffusion Processes and Fertility Transition*, ed. J. B. Casterline, Washington: National Academy Press, 66–114.
Papineau, D. (2001) "Metaphysics over Methodology – or, Why Infidelity Provides No Grounds to Divorce Causes from Probabilities", *Stochastic Causality*, ed. M. C. Galavotti, P. Suppes and D. Costantini, Stanford: Center for the Study of Language and Information, 15–38.
Pearl, J. (1998) "Graphs, Causality, and Structural Equation Models", *Sociological Methods & Research* 27, 226–84.
Pearl, J. (1999) "Graphs, Structural Models, and Causality", *Computation, Causation, and Discovery*, ed. C. Glymour and G. F. Cooper, Cambridge: MIT Press, 95–138.
Pearl, J. (2000) *Causality*, Cambridge: Cambridge University Press.
Pötter, U. (2006) "Probabilistische Selektionsmodelle", *Methoden der Sozialforschung*, ed. A. Diekmann, Wiesbaden: VS Verlag für Sozialwissenschaften, 172–202.
Pötter, U. and Blossfeld, H.-P. (2001) "Causal Inference from Series of Events", *European Sociological Review* 17, 21–32.
Przeworski, A. and Teune, H. (1970) *The Logic of Comparative Social Inquiry*, New York: Wiley.
Robins, J. M. (1999) "Association, Causation, and Marginal Structural Models", *Synthese* 121, 151–79.
Rogers, E. M. (1995) *Diffusion of Innovations*, 4th edn, New York: Free Press.
Rohwer, G. and Pötter, U. (2001) *Grundzüge der sozialwissenschaftlichen Statistik*, Weinheim: Juventa.
Rohwer, G. and Pötter, U. (2002a) *Methoden sozialwissenschaftlicher Datenkonstruktion*, Weinheim: Juventa.
Rohwer, G. and 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, 41–8.
Rosenbaum, P. R. (2002) *Observational Studies*, 2nd edn, New York: Springer-Verlag.
Rosenbaum, P. R. and Rubin, D. B. (1983) "The Central Role of the Propensity Score in Observational Studies for Causal Effects", *Biometrika* 70, 41–55.
Rothman, K. J. and Greenland, S. (2005) "Causation and Causal Inference in Epidemiology", *American Journal of Public Health* 95, Supp. 1, 144–50.
Rubin, D. B. (1974) "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies", *Journal of Educational Psychology* 66, 688–701.
Rubin, D. B. (1990) "Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies", *Statistical Science* 5, 472–80.
Ryle, G. (1949) *The Concept of Mind*, London: Hutchinson.

Scheines, R. (1997) "An Introduction to Causal Inference", *Causality in Crisis? Statistical Methods and the Search for Causal Knowledge in the Social Sciences*, ed. V. R. McKim and S. P. Turner, Notre Dame: University of Notre Dame Press, 185–99.

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. and 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", *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, ed. G. Arminger, C. C. Clogg and M. E. Sobel, New York: Plenum Press, 1–38.

Sobel, M. E. (2005) "Discussion: The Scientific Model of Causality", *Sociological Methodology* 35, 99–133.

Sober, E. (1982) "Frequency-Dependent Causation", *Journal of Philosophy* 79, 247–53.

Sober, E. (1986) "Causal Factors, Causal Inference, Causal Explanation", *The Aristotelian Society* 60, Supp., 97–113.

Spirtes, P., Glymour, C. and Scheines, R. (1993) *Causation, Prediction, and Search*, New York: Springer-Verlag.

Splawa-Neyman, J. (1990) "On the Application of Probability Theory to Agricultural Experiments", *Statistical Science* 5, 465–80.

Steyer, R., Gabler, S., Davier, A. A. von, Nachtigall, C. and Buhl, T. (2000) "Causal Regression Models I: Individual and Average Causal Effects", *Methods of Psychological Research Online* 5, No. 2, 39–71.

Strang, D. (1991) "Adding Social Structure to Diffusion Models", *Sociological Methods & Research* 19, 324–53.

Strang, D. and Tuma, N. B. (1993) "Spatial and Temporal Heterogeneity in Diffusion", *American Journal of Sociology* 99, 614–39.

Suppes, P. (1970) *A Probabilistic Theory of Causality*, Amsterdam: North-Holland.

Susser, M. (1991) "What is a Cause and How do we Know One? A Grammar for Pragmatic Epidemiology", *American Journal of Epidemiology* 133, 635–48.

Tryfos, P. (2004) *The Measurement of Economic Relationships*, Dordrecht: Kluwer.

Twardy, C. R. and Korb, K. B. (2004) "A Criterion of Probabilistic Causation", *Philosophy of Science* 71, 241–62.

Wasserman, S. and Faust, K. (1994) *Social Network Analysis: Methods and Applications*, Cambridge: Cambridge University Press.

Wellman, B. (1988) "Structural Analysis: From Method and Metaphor to Theory and Substance", *Social Structures: A Network Approach*, ed. B. Wellman and S. D. Berkowitz, Cambridge: Cambridge University Press, 19–61.

Winship, C. and Morgan, S. L. (1999) "The Estimation of Causal Effects from Observational Data", *Annual Review of Sociology* 25, 659–706.

Woodward, J. (1999) "Causal Interpretation in Systems of Equations", *Synthese* 121, 199–247.

Woodward, J. (2001) "Probabilistic Causality, Direct Causes and Counterfactual Dependence", *Stochastic Causality*, ed. M. C. Galavotti, P. Suppes and D. Costantini, Stanford: Center for the Study of Language and Information, 39–63.

Woodward, J. (2003) *Making Things Happen. A Theory of Causal Explanation*, Oxford: Oxford University Press.

Wooldridge, J. M. (2002) *Econometric Analysis of Cross Section and Panel Data*, Cambridge: MIT Press.

Yamaguchi, K. (1994) “Some Accelerated Failure-Time Regression Models Derived from Diffusion Process Models: An Application to a Network Diffusion Analysis”, *Sociological Methodology* 24, 267–300.

Index

actor’s rules 29
actors: substantial and secondary 84
aleatoric probability 33

Barker, S. 108
Bartholomew, D. J. 137
Bates, F. L. 131
Berkowitz, S. D. 131
Blalock, H. M. 137
Blau, P. M. 18, 19, 131
Blossfeld, H.-P. 136, 137
Bollen, K. A. 134
Borgatti, S. P. 131
Boyd, J. P. 131
Brady, H. E. 135
Brüderl, J. 137
Burt, R. S. 131
Buskens, V. 137

Campbell, D. T. 135
Carroll, J. W. 134
Cartwright, N. 72
causal effect: generic 64;
individual 62
Clogg, C. C. 133
closed stochastic models 46, 135, 136
cohort approach 32
Coleman, J. S. 21, 137
Collier, D. 135
comparative cause 61
concurrent causes 107
conditional distribution 7; temporal
form 97
conditional independence 43
constraints 43
context-dependency 58
contextual variables 95, 102
Cook, T. D. 135
Cooper, G. F. 133
covariate context 57
Cox, D. R. 134, 135

data generating process 20, 21, 82, 87
data model 71; statistical 71; stochastic 71
Dawid, A. P. 134, 136
dependence: of modal variables 42;
on context 58
descriptive generalization 37
descriptive regression model 9, 10
Diekmann, A. 137
diffusion models 120
distribution-dependent: causation 123;
effects 66; regression functions 114
Duncan, G. J. 137
dynamic cause 61
dynamic effect 104

Eells, E. 134
effective condition 42
elementary function 41
Elwert, F. 77
equivalence: automorphic 23; of
functional models 41; statistical 23;
structural 23
event data 96
event models 96
event types 94
event variables 95
events 28, 94
Everett, M. G. 131
experiment: hypothetical 83;
randomized 74

Faust, K. 131
Fisher, R. A. 6
Freedman, D. A. 131, 133
frequency function 4
frequency-dependent causation 137

functional cause: comparative 61; definition 57, 59: dynamic 61
functional effect 57
functional models 37, 38; as process frames 40; equivalence, 41; pseudo-indeterministic 51; structural identity 40, 46
functions: deterministic 44; stochastic 45

generalization: descriptive 37; modal 37
generic situation 95
generic unit 23
Glenn, N. D. 32
Glymour, C. 133
Goldthorpe, J. H. 134, 136
Greene, W. H. 131
Greenland, S. 133
Greve, H. R. 137

Haritou, A. 133
Hausman, D. M. 133, 136
Heckman, J. J. 134–136
Hernes, G. 22
Hitchcock, C. R. 134
Holland, P. W. 61, 79, 134, 135
Homans, G. C. 12, 17
homogeneity assumptions 64
hypothetical experiment 83

independence: conditional 43; of modal variables 42
individual-level models 111
interaction 14
interactive conditions 66, 89
intervening causes 106
interventions 83; modal 84; substantial 84

Jauss, H.-R. 131

Keohane, R. O. 134
King, G. 134
Knoke, D. 131
Korb, K. B. 134
Kuklinski, J. H. 131

McKim, V. R. 9, 134
macro facts 21
Maddala, G. S. 88, 136
Mahajan, V. 137
Marini, M. M. 134
mean value 3
mean value regression 8
micro facts 20
Mill, J. S. 133, 134
modal interventions 84
modal questions 63; and counterfactual questions 81, 134
modal variables 38, 39, 113; endogenous 39; exogenous 39
modal view of relations 15
modeling rules 29, 39
Morgan, S. L. 134
Morris, M. 137
multi-dimensional variables 43
multilevel models 115, 117

networks 12
Neyman, J. 135
non-occurring events 101
Nurmi, H. 136

Oakeshott, M. 131
observational data 81

Palloni, A. 137
Papineau, D. 133
Peacock, W. G. 131
Pearl, J. 133, 135, 136
Peterson, R. A. 137
Pötter, U. 136, 137
population-level models 74, 111; deterministic 113; stochastic 116
population-level variables: deterministic 113; stochastic, 116
populations and reference sets 2
potential outcome approach 61, 79, 90
praxeological questions 84
prediction functions 48
prediction with functional models 43
principle of common dependence 44
probability distribution of a stochastic variable 34
process frames 28; and models 32, 40; stochastic 35; timed 30
processes: and events 28; and rules 29; as constructions 27; data generating 82; historical 27; mechanical 132; repeatable, 28; substantial 82
property space 2, 23; numerical representation 3
Przeworski, A. 131
pseudo-indeterministic models 51, 54

random generators 32
random variables 34
randomization 75; as substantial process 82, 87, 91

randomized experiment 74
range 2, 34
Raudenbush, S. W. 137
reduced model 42, 48
reference set 2
regression functions:
distribution-dependent 114; general 7; specific 8
regression models 9, 51
relational variables 11
relations: comparative 14; context-dependent 14; interactive 14; modal view 15
residuals 54
risk function 100
Robins, J. M. 133
Rogers, E. M. 137
Rosenbaum, P. R. 62, 81, 135
Rothman, K. J. 133
Rubin, D. B. 61, 135
Ryle, G. 29

Scheines, R. 133, 136
Seawright, J. 135
secondary actors 84
self selection 87; with constraints 126
set-valued functions 42
Sewell, W. H. 16
Shadish, W. R. 135
Singer, B. 134
Smith, B. C. 132
Sobel, M. E. 134, 136
Sober, E. 134, 137
social structure 17; statistical notion 18
Spirtes, P. 133
state variables 96
statistical approach 4, 23, 63, 64, 113; and network approach 24
statistical descriptions 5
statistical distributions 4; as functions 4
statistical processes 30; synchronously aggregated 31; temporally aggregated 31
Steyer, R. 134
stochastic effect 59
stochastic equivalence of functional models 48
stochastic variables 34
Strang, D. 137
structural conditions 22
structural equations models 52
structural identity of functional models 40, 46
structure: relational notion 22; statistical notion 17
structured units 22, 24, 113
substantial actors 84
substantial interventions 84
substantial processes 21, 82, 84, 88, 126
summarized risk function 100
Suppes, P. 134
survivor function 100
Susser, M. 133

temporal locations 30
Teune, H. 131
time axis: continuous 30; discrete 30
time series 30
time-dependent effects 109
timed process frames 30
Tryfos, P. 133
Tuma, N. B. 137
Twardy, C. R. 134

unobserved variables 52

variable: logical 2; statistical 2
Verba, S. 134

Wasserman, S. 131
Wellman, B. 131
Wermuth, N. 135
Winship, C. 77, 134
Woodward, J. 83, 86, 133, 135, 136
Wooldridge, J. M. 134, 136

Yamaguchi, K. 137