Multilevel and Population-level Models

G. Rohwer

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 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 parent's 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 parent's 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 to take 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. 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, and discusses how this notion can be used for multilevel and population-level models.

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

[1] See the critical remarks made by Coleman (1990: 1).

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:

$$\ddot{X}^* \longrightarrow \ddot{Y}^* \tag{1}$$

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. 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.

[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. Distribution-dependent Regression Functions. The model in (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 a regression function

$$x \longrightarrow \mathrm{P}[Y \mid X = x]$$

can 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

$$(x, \mathrm{P}[X]) \longrightarrow \mathrm{P}[Y \mid X = x, \ddot{X}^* = \mathrm{P}[X]] \tag{2}$$

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

$$\mathrm{P}(Y = 1 \mid X = j, \ddot{X}^* = \mathrm{P}[X]) = \min\{1, s_j/(n\,\mathrm{P}(X = j))\} \tag{3}$$

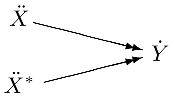
4. Corresponding Individual-level Models? If a population-level model results from independent repetitions of an individual-level model, as was assumed in § 1, it obviously leads to distribution-independent regression functions. On the other hand, starting from a population-level model that implies distribution-dependent 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 (2)) has the following form:

$$(x, \mathrm{P}[X]) \longrightarrow \Pr[\dot{Y} \mid \ddot{X} = x, \ddot{X}^* = \mathrm{P}[X]] \tag{4}$$

The corresponding functional model may be depicted as



(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 (3), one would get the formulation

$$\Pr(\dot{Y} = 1 \mid \ddot{X} = j, \ddot{X}^* = \mathrm{P}[X]) = \min\{1, s_j/(n\,\mathrm{P}(X = j))\} \tag{6}$$

In contrast to the deterministic population-level model (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 (5) are therefore different from population-level models. Their focus on an individual-level endogenous variable suggests to think 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 (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 $\ddot{X} \longrightarrow\!\!\!\!\rightarrow \dot{Y}$. 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

$$\Pr(\dot{Y} = 1 \mid \ddot{X} = 0) = \pi_{01} \quad \text{and} \quad \Pr(\dot{Y} = 1 \mid \ddot{X} = 1) = \pi_{11}$$

Let now $\mathrm{P}[X] \in \mathcal{D}[\Omega, \tilde{\mathcal{X}}]$ be the distribution of a statistical variable X, representing the parents' educational levels. The numbers 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

$$\ddot{X}^* \longrightarrow\!\!\!\!\rightarrow \dot{Y}^* \tag{7}$$

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

$$\mathrm{P}[X] \longrightarrow \Pr(\dot{Y}^* = \mathrm{P}[Y] \,|\, \ddot{X}^* = \mathrm{P}[X]) \tag{8}$$

To each value 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 (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].

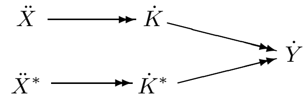
This then allows one to derive a multilevel model in the sense of § 5. The model includes a deterministic population-level variable $\ddot{X}^*$, a correspondingly defined deterministic individual-level variable $\ddot{X}$, and a stochastic individual-level variable $\dot{Y}$. The distribution of $\dot{Y}$ 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 parent's 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.



(9)

Three processes can be distinguished.

a) The individual-level process $\ddot{X} \longrightarrow\!\!\!\!\rightarrow \dot{K}$ 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 $\ddot{X}^* \longrightarrow\!\!\!\!\rightarrow \dot{K}^*$ 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

$$\Pr(\dot{Y} = 1 \,|\, \ddot{X} = j, \ddot{X}^* = \mathrm{P}[X], \dot{K}^* = k) = \pi_j \, \min\{1, \frac{s}{k}\}$$

Assuming that $\dot{K}^*$, given $\ddot{X}^*$, is (approximately) independent of $\ddot{X}$, it is

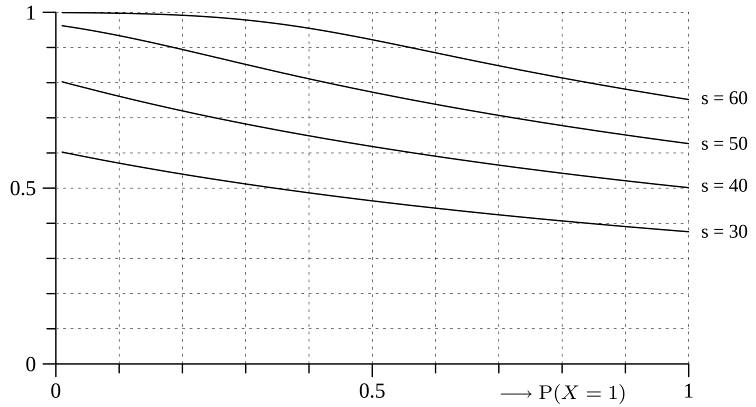


Figure 1 Dependence of the sum term in (10) on $\mathrm{P}(X = 1)$ for different values of s. Parameters: $n = 100$, $\pi_0 = 0.5$, and $\pi_1 = 0.8$.

possible to average over possible outcomes:

$$\Pr(\dot{Y} = 1 \mid \ddot{X} = j, \ddot{X}^* = \mathrm{P}[X]) = \pi_j \sum_{k=0}^{n} \min\{1, \frac{s}{k}\} \Pr(\dot{K}^* = k \mid \ddot{X}^* = \mathrm{P}[X]) \tag{10}$$

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 $n = 100$, $\pi_0 = 0.5$, and $\pi_1 = 0.8$, Figure 1 illustrates the dependence of this mean proportion on $\mathrm{P}(X = 1)$ for different values of s. Multiplication with π_j would show how the probability of being accepted depends on the distribution of X.

Thus, (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.

2 Models of Statistical Processes

1. Process Frames and Models. Statistical processes have been defined as temporal sequences of statistical variables. They may be depicted as

$$Y_0 \longrightarrow Y_1 \longrightarrow Y_2 \longrightarrow Y_3 \longrightarrow \cdots \tag{11}$$

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

$$\ddot{Y}_0^* \longrightarrow\!\!\!\!\!\rightarrow \dot{Y}_1^* \longrightarrow\!\!\!\!\!\rightarrow \dot{Y}_2^* \longrightarrow\!\!\!\!\!\rightarrow \dot{Y}_3^* \longrightarrow\!\!\!\!\!\rightarrow \cdots$$

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 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

$$\begin{array}{l} \dot{Y}_t \searrow \\ \qquad\quad \dot{Y}_{t+1} \\ \dot{N}_t^* \nearrow \end{array} \tag{12}$$

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 already are in the specified state. A simple specification of the stochastic function is

$$\mathrm{P}(\dot{Y}_{t+1} = 1 \mid \dot{Y}_t = 0, \dot{N}_t^* = n_t) = \alpha \frac{n_t}{n} \tag{13}$$

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

$$\Pr(\dot{N}_{t+1}^* = n_{t+1} \mid \dot{N}_t^* = n_t) = \binom{n - n_t}{n_{t+1} - n_t} \pi_t^{n_{t+1} - n_t} (1 - \pi_t)^{n - n_{t+1}} \tag{14}$$

[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.

Box 1 Simulation of diffusion processes according to (13) and (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.

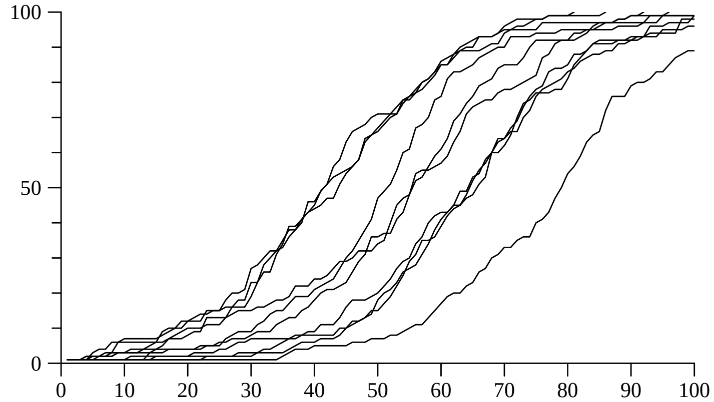


Figure 2 Ten simulated diffusion processes, generated with the algorithm shown in Box 1. Parameters: $n = 100$, $n_0 = 1$, $\alpha = 0.1$.

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 2 shows ten simulated diffusion paths, generated with the algorithm in Box 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 (13) and (14).

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 $r(t) := \Pr(\dot{Y}_{t+1} = 1 \,|\, \dot{Y}_t = 0)$. This allows one to derive a

[5] See, e.g., Diekmann (1989), Brüderl and Diekmann (1995).

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,

$$\mathcal{N}_t^0 := \{\omega \,|\, Y_t(\omega) = 0\} \quad \text{and} \quad \mathcal{N}_t^1 := \{\omega \,|\, Y_t(\omega) = 1\}$$

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 another idea 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

[6] This is assumed in the approaches proposed by Strang (1991), Strang and Tuma (1993), Greve, Tuma and Strang (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).

$\tilde{x}_l$. Given then a statistical variable $X : \Omega \longrightarrow \tilde{\mathcal{X}}$, one can define groups $\Omega_j := \{\omega \,|\, X(\omega) = \tilde{x}_j\}$ $(j = 1, \ldots, m)$, 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 (13) into

$$\Pr(\dot{Y}_{t+1} = 1 \,|\, \dot{Y}_t = 0, \ddot{X} = \tilde{x}_j, \dot{N}_{1t}^* = n_{1t}, \ldots, \dot{N}_{mt}^* = n_{mt}) = \sum_{k=1}^{m} \delta_{jk}\, \alpha_k\, \frac{n_{kt}}{n} \tag{15}$$

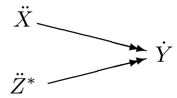
This is now a multilevel diffusion model where the probability of getting the specified property depends on the group, Ω_j, an individual unit belongs to and, for $k = 1, \ldots, m$, on the value of $\dot{N}_{kt}^*$ and the proximity between Ω_j and Ω_k. To complete the model, formally analogous to (14), one can set up a separate equation for each group:

$$\Pr(\dot{N}_{j,t+1}^* = n_{j,t+1} \,|\, \dot{N}_{1t}^* = n_{1t}, \ldots, \dot{N}_{mt}^* = n_{mt}) = \binom{n_j - n_{jt}}{n_{j,t+1} - n_{jt}} \pi_{jt}^{n_{j,t+1} - n_{jt}} (1 - \pi_{jt})^{n_j - n_{j,t+1}} \tag{16}$$

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 1. (Illustrations will be discussed in the next section.)

3 Functional Causality and Levels

1. Distribution-dependent Causation. Assume a multilevel model as defined in § 5 of Section 1 that has the form



(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 $\mathrm{P}[Z]$. 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 $\mathrm{P}[Z]$ changes the relationship between $\ddot{X}$ and $\dot{Y}$.

The question can be considered in two forms. The example of § 8 in Section 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 (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 1 by multiplying the curves with π_j. The solid line in Figure 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. The converse holds for children having parents with 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 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.

[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).

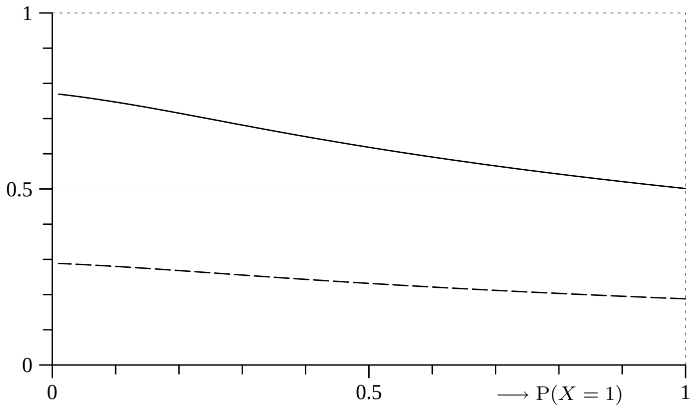


Figure 3 Dependence of $\Pr(\dot{Y}=1|\ddot{X}=1,\ddot{X}^*=\mathrm{P}[X])$ (solid line) and $\mathrm{E}(\dot{Y}|\ddot{X}=1,\ddot{X}^*=\mathrm{P}[X])-\mathrm{E}(\dot{Y}|\ddot{X}=0,\ddot{X}^*=\mathrm{P}[X])$ (broken line) on $\mathrm{P}(X=1)$. Parameters: $n=100$, $s=50$, $\pi_0=0.5$, and $\pi_1=0.8$.

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:

$$\begin{aligned}\mathrm{P}(Y=1) \approx\ & \Pr(\dot{Y}=1\,|\,\ddot{X}=0,\ddot{X}^*=\mathrm{P}[X])\,\mathrm{P}(X=0)+ \\ & \Pr(\dot{Y}=1\,|\,\ddot{X}=1,\ddot{X}^*=\mathrm{P}[X])\,\mathrm{P}(X=1)\end{aligned} \tag{18}$$

Figure 4 shows how this proportion depends on $\mathrm{P}(X=1)$, 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

[9]This does not exclude the possibility that selecting the value of an individual-level

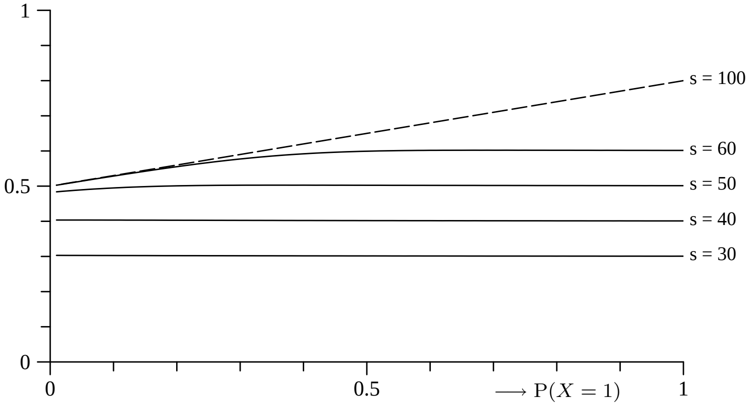
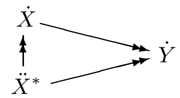


Figure 4 Dependence of $\mathrm{P}(Y=1)$ as defined in (18) on $\mathrm{P}(X=1)$ for selected values of s. Parameters: $n=100$, $\pi_0=0.5$, and $\pi_1=0.8$.

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, $\mathrm{P}[X]$, 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 (5) in Section 1. Would it make sense to change $\ddot{X}$ into an endogenous variable $\dot{X}$ resulting in the model



(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

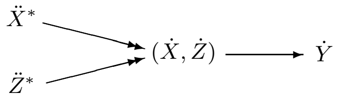
variable may constrain the range of possible values of a (corresponding) population-level variable.

[10]This has been stressed in the literature dealing with modeling contextual effects, see, e.g., Blalock (1984), Duncan and Raudenbush (1999).

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: The educational success $\dot{Y}$ (0 or 1) depends on the parents' educational level $\ddot{X}$ (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 $\ddot{X}$.

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:



(20)

Of course, one needs assumptions about the selection processes that generate the distribution of $(\dot{X}, \dot{Z})$. Here we continue with a previous numerical illustration 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:

$$\Pr(\dot{X} = 1, \dot{Z} = 2 \,|\, \ddot{X}^* = \mathrm{P}[X], \ddot{Z}^* = \mathrm{P}[Z]) = \min\{\Pr(\dot{Z} = 2 \,|\, \dot{X} = 1)\, \mathrm{P}(X = 1), \mathrm{P}(Z = 2)\} \qquad (21)$$

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, $\mathrm{E}_1 - \mathrm{E}_0$ with

$$\mathrm{E}_j := \mathrm{E}(\dot{Y} | \dot{X} = j, \ddot{X}^* = \mathrm{P}[X], \ddot{Z}^* = \mathrm{P}[Z]) \qquad (22)$$

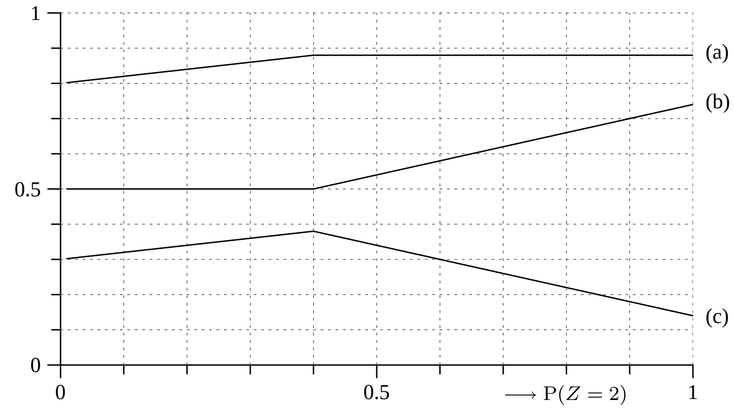


Figure 5 Dependence of (a) E_1, (b) E_0, and (c) $\mathrm{E}_1 - \mathrm{E}_0$, as defined in (22), on $\mathrm{P}(Z = 2)$ (proportion of schools of type 2).

depends on the distribution of school types. Based on the numerical specification of the function $(\dot{X}, \dot{Z}) \longrightarrow\!\!\!\!\rightarrow \dot{Y}$ previously assumed, this is shown in Figure 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 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

$$\alpha_1 = 0.2,\ \alpha_2 = 0.1,\ \delta_{11} = 1,\ \delta_{12} = \delta_{21} = 0.1,\ \delta_{22} = 0.5,$$

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 1. Figure 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 (15). Figure 7 illustrates their development for the example. These are now time-dependent effects. Moreover, the effects heavily depend on the distribution of X. Figure 7 was generated with 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.

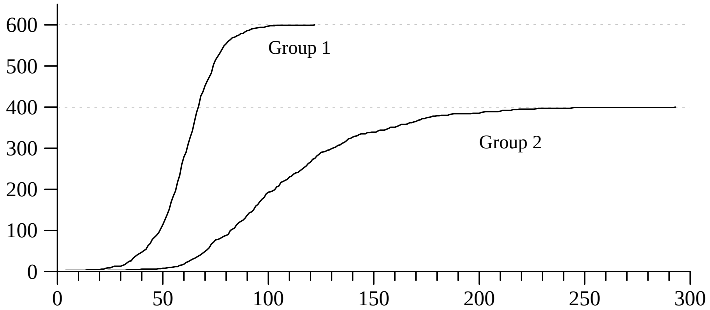


Figure 6 Simulation of a structured diffusion processes with two groups. Parameter values are defined at the beginning of § 4. Initial values are $n_{1,0} = n_{2,0} = 1$.

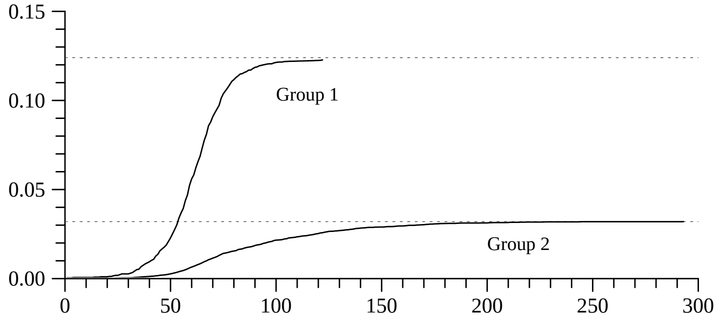


Figure 7 Development of the conditional probabilities defined in (15), simulated with group sizes $n_1 = 600$ and $n_2 = 400$.

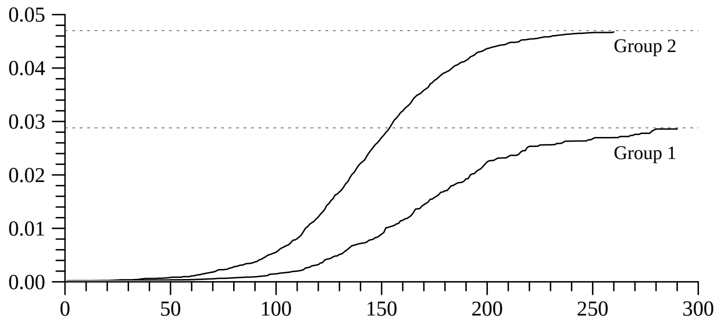


Figure 8 Development of the conditional probabilities defined in (15), simulated with group sizes $n_1 = 100$ and $n_2 = 900$.

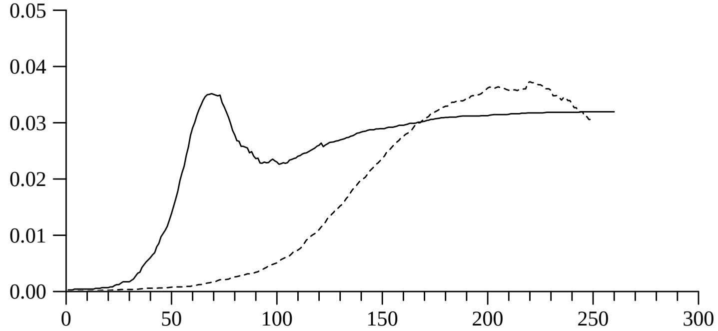


Figure 9 Development of the overall rates defined in (23). Solid line: $n_1 = 600$ and $n_2 = 400$, dotted line: $n_1 = 100$ and $n_2 = 900$.

7. Mixing Group-level Effects. The previous paragraph considered the conditional probabilities defined in (15) separately for each group. Overall rates, for the whole population, can be derived by mixing the group-specific rates in the following way:

$$\Pr(\dot{Y}_{t+1} = 1 \,|\, \dot{Y}_t = 0, \dot{N}^*_{1t} = n_{1t}, \ldots, \dot{N}^*_{mt} = n_{mt}) = \sum_j \Pr(\dot{Y}_{t+1} = 1 \,|\, \dot{Y}_t = 0, \ddot{X} = \tilde{x}_j, \ldots)\, \mathrm{P}(X = \tilde{x}_j \,|\, \dot{Y}_t = 0) \qquad (23)$$

Mixing is with proportions $\mathrm{P}(X = \tilde{x}_j \,|\, \dot{Y}_t = 0)$, denoting the proportion of members of group $\tilde{x}_j$ in the risk set at t. Figure 9 illustrates these overall rates for two different distributions of the group variable X.