Basic Notions of Event History Analysis

G. Rohwer

U. Pötter

July 2009

Contents

Chapter 1

Introduction

We discuss statistical methods for longitudinal data from a life course perspective. This chapter briefly introduces a few elementary notions.

1.1 Life Courses

We consider human individuals, or other objects, which exist in the form of a life course: The life course begins with the individual's birth and ends with its death.

A life course will be considered as a sequence of states. The notion presupposes, and is dependent on, a *state space*, that is, a predefined set of possible states. The state space will be denoted by $\mathcal{Y}$.

1.2 Biographical Frames

The specification of a set of possible life courses will be called a *biographical frame*. It can be represented by a directed graph where the nodes represent states and the arcs represent possible transitions.

Exercise 1.1 Specify a biographical frame based on a state space comprising the states (1) unemployed, (2) employed, (3) out of labor force, (4) dead.

1.3 Multidimensional State Spaces

A multidimensional state space consists of two or more, say m, state spaces combined in the following way:

$$\mathcal{Y} = \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_m$$

Exercise 1.2 Define a state space comprising the states (1) not married, (2) married. Then create a two-dimensional state space that also contains the states introduced in Exercise 1.1.

Exercise 1.3 Specify a biographical frame based on the two-dimensional state space defined in Exercise 1.2.

1.4 Time Axes

The basic idea is to think of a life course as a sequential walk through a state space. This obviously requires the reference to a time axis. There are two possibilities:

Box 1.1 Data set 1

ID	Birth	Begin of Study	End of Study
1	1970	1990	1995
2	1975	1994	1999
3	1973	1991	1996
4	1970	1989	1995
5	1975	1993	1999
6	1973	1993	1996
7	1970	1988	1995
8	1975	1995	1999
9	1973	1992	1997

- One can use a *discrete time axis*, defined as a sequence of temporal locations (e.g., days, months, years), numerically represented by (a subset of) the set of integers.
- Or one can use a *continuous time axis*, numerically represented by (a subset of) the set of real numbers.

We begin with assuming a discrete time axis. This allows to think of a *sequence* of temporal locations.

1.5 Events

We will use a narrow notion of events, defined as changes from one state into another state.

Exercise 1.4 Describe all events which are possible in the biographical frame developed in Exercise 1.1.

1.6 Calendar Time and Process Time

While the collection of data must begin with a calendar time axis, the description and modeling of life courses most often employs a *process time axis*, that is, a time axis that begins with a specified event, e.g., the birth of an individual, or the beginning of a partnership.

Exercise 1.5 Consider the data in Box 1.1. Suppose the data refer to students and the dates they start and finish their studies. Define a state space and a biographical frame. Present the data on a process time axis that starts from the beginning of study.

Chapter 2

Statistical Descriptions

How to describe life courses? There are two complementary approaches. One can consider life courses of particular individuals and provide descriptions of their specific development; or one can start from a (relatively large) set of life courses and develop statistical descriptions. Here we follow a statistical approach.

It will be assumed that one can refer to a collection of individuals, denoted by Ω, and for each individual there is a life course (based on the same biographical frame). In this chapter, we briefly introduce a few statistical notions that will be used in subsequent chapters.

2.1 Statistical Variables

A statistical variable will be defined as a function

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

To each individual $\omega \in \Omega$, the variable X assigns a value $X(\omega)$ that is an element of the property space $\mathcal{X}$. It will be assumed that the property space has a numerical representation and can be viewed as a set of real numbers. The variable is called *discrete* if its property space can be represented by a subset of the natural numbers. Otherwise it is called *continuous*.

2.2 State Variables

Given a biographical frame based on a state space $\mathcal{Y}$, statistical variables can be used to represent life courses. We first assume a discrete time axis

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

One can define state variables

$$Y_t : \Omega \longrightarrow \mathcal{Y}$$

$Y_t(\omega)$ is the state of the individual $\omega \in \Omega$ in the temporal location t. Each individual life course is then given by a sequence

$$(Y_0(\omega), Y_1(\omega), Y_2(\omega), \ldots)$$

2.3 Partial Life Courses

Social research is most often not concerned with complete life courses that begin with birth and end with the death of an individual. The question then occurs how to define partial life courses. There are two possibilities. In both cases, the partial life courses begin with a specified type of event, e.g., the beginning of a job or marriage.

a) The first possibility is to fix a maximum duration, e.g., until age 20, or the first 10 years after the beginning of a new job. Formally, one specifies a maximum duration t^*, and partial life courses are then given by sequences

$$(Y_0(\omega), Y_1(\omega), Y_2(\omega), \ldots, Y_{t^*}(\omega))$$

Of course, realized life courses might be shorter than t^*.

b) Another possibility is to define the end of a partial life course when a specific event occurred, e.g., an individual's death, or the end of a job.

Social research most often uses the second approach.

2.4 Statistical Distributions

Statistical descriptions are based on distributions of statistical variables. The basic idea is that one is not interested in the attributes of particular individuals, but in frequencies of attributes defined for a collection of individuals. This has been well formulated in the "Declaration on Professional Ethics" published by the *International Statistical Institute*:

> "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 cooperating (or non-cooperating) subjects should therefore be kept confidential, whether or not confidentiality has been explicitly pledged."[1]

A *statistical distribution* is defined as a function

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

$\mathcal{A}(\mathcal{X})$ is a set of subsets of the property space $\mathcal{X}$ (assumed to be complete with respect to set-theoretic operations). The elements of $\mathcal{A}(\mathcal{X})$ are called *property sets*. The function P assigns to each property set $\tilde{x} \in \mathcal{A}(\mathcal{X})$ the proportion of members of Ω having values in this set, i.e.:

$$\mathrm{P}(\tilde{x}) := |\, \{\omega \in \Omega \,|\, X(\omega) \in \tilde{x}\} \,| \,/\, |\, \Omega \,|$$

[1] International Statistical Institute 1986, p. 238.

Box 2.1 Data set 2

```
ID   t = 0  1  2  3  4  5
-------------------------
 1       0  0  1  1  0  1
 2       1  0  0  0  1  1
 3       1  1  0  0  0  0
 4       0  0  0  1  1  1
 5       0  1  1  1  0  1
 6       1  1  0  0  1  1
```

Using this definition, there is no longer any reference to identifiable individuals; instead, the only information is about frequencies of properties in the collection of individuals.

For easy reference to property sets, we shall also use the following notations:

$$\mathrm{P}(X \in \tilde{x}) := \mathrm{P}(\tilde{x})$$

$$\mathrm{P}(X = x) := \mathrm{P}(\{x\})$$

A further notation for quantitative variables is

$$\mathrm{P}(X \leq x) := \mathrm{P}(\{\omega \in \Omega \,|\, X(\omega) \leq x\})$$

called the *distribution function* of the variable X. A commonly used abbreviation is $F(x) := \mathrm{P}(X \leq x)$.

Exercise 2.1 Assume that Ω is a collection of 10 individuals and there are the following values of a variable X:

$$3, 2, 3, 1, 4, 3, 1, 3, 4, 2$$

(a) Specify a minimal property space of X. (b) Calculate the distribution function of X.

2.5 State Distributions

A simple approach to the statistical representation of life courses uses state variables. These statistical variables represent for each point in time the states occupied by the individuals (see Section 2.2).

Exercise 2.2 Consider the data in Box 2.1, interpreted as job histories of six individuals (1 = employed, 0 = unemployed). Calculate the state distributions and provide a graphical illustration.

Sequences of state distributions are most informative if states cannot be repeated. For example: 0 = never been married, 1 = married, or having been married. However, if states are repeatable, sequences of state distributions can easily be misleading.

Exercise 2.3 Construct an example that illustrates the problem. Assume two repeatable states: 1 = employed, 0 = not employed. Now construct two variants of six processes such that the proportion of unemployed individuals is always 1/3. In the first variant, two individuals are always, and four individuals are never unemployed; in the second variant all individuals have identical unemployment durations.

Chapter 3

Duration Distributions

This chapter introduces some statistical notions describing episode data. Throughout the chapter, it is assumed that observations are complete. How to use incomplete (right censored) observations will be discussed in the next chapter.

3.1 Episodes

Given a biographical frame, a life course shows how an individual sequentially stays in the states of the state space. The individual starts in some specific state, remains in that state for some time, then changes into another state and remains in that state for some time, and so on until the (observed) life course ends. This suggests to view a life course as a sequence of episodes, each episode being characterized by four pieces of information:

- an *origin state* (a change into this state indicates the beginning of the episode);
- a *destination state* (a change into this state indicates the end of the episode);
- a *starting time* (at which the origin state takes place for the first time);
- an *ending time* (at which the destination state takes place for the first time).

This notion suggests a general scheme to represent episode data. It will be called an *episode data scheme*. Box 3.1 provides an illustration with four life courses. The state space consists of four states. Each row represents one episode of an individual. The abbreviations have the following meaning:

- `ID` identifies the individuals,
- `SN` counts the individual's episodes,
- `ORG` is the origin state,
- `DES` is the destination state,
- `TS` is the starting time,
- `TF` is the ending time.

Box 3.1 A general scheme for episode data (data set 3)

```
ID   SN   ORG   DES  TS   TF
----------------------------
 1    1     1     2   0   10
 1    2     2     3  10   15
 1    3     3     4  15   20
 2    1     1     4   0   15
 3    1     1     3   0   16
 3    2     3     4  16   18
 4    1     1     2   0    6
 4    2     2     3   6   11
 4    3     3     2  11   17
 4    4     2     4  17   23
```

Exercise 3.1 Using the data in Box 1.1, first construct a biographical frame, and then present the data in the form of an episode data scheme.

Exercise 3.2 Using the data in Box 2.1, first construct a biographical frame, and then present the data in the form of an episode data scheme.

3.2 Statistical Framework

We now consider an approach to the description of life courses that takes episodes as its starting point. The idea is to refer to the collection of all episodes that begin in the same origin state. These episodes are then compared on a common process time axis where each episode begins at time zero. Such a collection of episodes can be represented by a two-dimensional statistical variable

$$(T, D)$$

where T denotes the duration and D the destination state of the episodes, respectively.

3.3 Single Destination State

If there is only a single destination state, it suffices to describe the distribution of the duration variable T. The conceptual tools depend on whether one uses a discrete or a continuous time axis. In both cases, the distribution function can be defined by

$$F(t) = \mathrm{P}(T \le t)$$

and the survivor function can be defined by

$$G(t) = \mathrm{P}(T \ge t)$$

Somewhat different definitions are used for the density and rate function. For a discrete time axis, the density function is defined by

$$f(t) = \mathrm{P}(T = t)$$

and the rate function is defined by

$$r(t) = \mathrm{P}(T = t \,|\, T \geq t)$$

For a continuous time axis, the density function is defined by

$$f(t) = \lim_{\Delta \to 0} \frac{\mathrm{P}(t \leq T < t + \Delta)}{\Delta}$$

and the rate function is defined by

$$r(t) = \lim_{\Delta \to 0} \frac{\mathrm{P}(t \leq T < t + \Delta \,|\, T \geq t)}{\Delta}$$

Exercise 3.3 Show, both for a discrete and for a continuous time axis, that the four notions just introduced are equivalent. In particular, derive the equation

$$r(t) = f(t)/G(t)$$

which is independent of the form of the time axis, the equation

$$G(t) = \prod_{\tau=0}^{t-1} (1 - r(\tau))$$

for a discrete time axis, and the equation

$$G(t) = \exp\left\{ - \int_0^t r(\tau)\, d\tau \right\}$$

for a continuous time axis.

Exercise 3.4 Using the data set 1 in Box 1.1, calculate the discrete rate function for study durations.

Exercise 3.5 Based on data set 2 in Box 2.1, consider the groups of episodes beginning, respectively, in state 0 and in state 1. For each group, only use episodes that have an identifiable destination state. Then calculate rate functions for the transition from 0 to 1, and from 1 to 0.

3.4 Multiple Destination States

If there are two or more possible destination states, one needs a two-dimensional variable (T, D) with D providing the destination state. Very helpful is then the notion of *state-specific rate functions*. Based on a discrete time axis, they are defined by

$$r_d(t) = \mathrm{P}(T = t, D = d \,|\, T \geq t)$$

where d indicates a destination state. Analogously, for a continuous time axis, the definition is

$$r_d(t) = \lim_{\Delta \to 0} \frac{\mathrm{P}(t \leq T < t + \Delta, D = d \,|\, T \geq t)}{\Delta}$$

The set of possible destination states will be denoted by $\mathcal{D}$, and we assume as a convention that

$$\mathcal{D} = \{1, \ldots, m\}$$

if there are m possible destination states.

Exercise 3.6 Consider in Box 3.1 all episodes which begin in state 1. Determine the set $\mathcal{D}$ of possible destination states, and for each $d \in \mathcal{D}$, calculate the rate function $r_d(t)$.

Exercise 3.7 Even if episodes can end in several destination states, one can ignore the distinctions and only consider a single destination state ('end of the episode'). Show that

$$r(t) = \sum_{d \in \mathcal{D}} r_d(t)$$

and illustrate this equation with the results of Exercise 3.6.

Exercise 3.8 Given episodes with multiple destination states on a continuous time axis, one can define so-called *sub-survivor functions*

$$G_d(t) = \exp\left\{ - \int_0^t r_d(\tau)\, d\tau \right\}$$

Show that the relationship

$$G(t) = \prod_{d \in \mathcal{D}} G_d(t)$$

holds. Then consider the question whether the sub-survivor functions can be given a sensible interpretation.

Chapter 4

Censored Observations

So far we assumed that observations are complete, i.e. the duration and the destination state of each episode are known. In practice, this is often not the case. This chapter considers the special case that some observation are right censored.

4.1 Right Censored Observations

The observation of an episode is called *right censored* if one only knows its duration up to some point in time, but neither the completed duration nor the destination state (if there are several possibilities). To formally represent this we start from a statistical variable (T, D) where the set of possible destination states is given by $\mathcal{D}$. Observations for $i = 1, \ldots, n$ individuals provide values not immediately of (T, D), but of another variable (T^*, D^*). D^* can take values in a set

$$\mathcal{D}^* = \mathcal{D} \cup \{0\}$$

where 0 represents the origin state of the episode and consequently is not a possible element of $\mathcal{D}$.[1] Observation are given as

$$(t_i^*, d_i^*) \qquad \text{for } \ i = 1, \ldots, n$$

and are connected with values (t_i, d_i) of the theoretically assumed variable (T, D) in the following way:

a) If $d_i^* \in \mathcal{D}$, the observation is complete: $t_i = t_i^*$ and $d_i = d_i^*$.

b) If $d_i^* = 0$, the observation is right censored and one only knows that $t_i > t_i^*$ (or $t_i \geq t_i^*$).

This allows to represent complete and right censored observations in the same formal framework (cf. Section 3.1). Right censored observations are indicated by the value zero in the destination state variable, implying that the value of the duration variable provides not the completed but the hitherto observed episode duration.

Exercise 4.1 Reorganize the data of Box 2.1 as episode data using the just mentioned convention to indicate right censored observations.

[1] Remember the convention to represent destination states by positive natural numbers. Therefore, $\mathcal{D}^* = \{0, 1, \ldots, m\}$ if there are m possible destination states.

Box 4.1 Data set 4

ID	DUR	CEN
1	17	1
2	5	0
3	22	1
4	13	1
5	2	0
6	9	1
7	12	0
8	15	1

4.2 Calculation of Survivor Functions

We begin with episodes having only a single destination state. Consequently, $\mathcal{D}^* = \{0, 1\}$, with 0 indicating censored and 1 indicating complete observations. The direct calculation of a survivor function is obviously not possible since one does not know all values t_i. However, one can calculate lower and upper limits for the unknown survivor function $G(t)$.

a) One gets a lower limit, denoted by $G^{\vdash}(t)$, if one assumes, for right censored observations, that the episode ends immediately after the observation period, i.e., at $t_i = t_i^*$ or $t_i = t_i^* + 1$.

b) One gets an upper limit, denoted by $G^{\dashv}(t)$, if one assumes that completed durations of right censored episodes are longer than the longest completely observed episode.

The unknown survivor function $G(t)$ is somewhere between these limits:

$$G^{\vdash}(t) \leq G(t) \leq G^{\dashv}(t)$$

Of course, the (time-varying) widths of the interval depend on the proportion of censored observations and on their distribution on the time axis.

Exercise 4.2 Using the data in Box 4.1, calculate lower and upper limits of the survivor function.

4.3 The Kaplan-Meier Procedure

If one doesn't know something exactly, one can try to estimate it. A procedure for estimating the survivor function $G(t)$ when the data contain right censored observations was proposed by E. L. Kaplan and P. Meier (1958). In order to explain the procedure, we first assume a discrete time axis. There is then the following relationship between the rate and the survivor function (see Section 3.3):

$$G(t) = \prod_{\tau=0}^{t-1} (1 - r(\tau))$$

This suggests to begin with estimates of the rate function $r(t)$, and then to use the formula to derive an estimate of the survivor function.

Obviously, if all observations were complete, one could calculate the rate function with

$$r(t) = \frac{E(t)}{R(t)}$$

where $E(t)$ denotes the number of episodes ending at t, and $R(t)$ denotes the number of episodes that ended not earlier than t. If some observations are right censored, one does not know these quantities, but it might be possible to estimate comparable quantities: $E^*(t)$, the number of episodes that are observed to end at t; and $R^*(t)$, the number of episodes which are not completed, or censored, earlier than t. With the help of these quantities one can define an observed rate function

$$r^*(t) = \frac{E^*(t)}{R^*(t)}$$

to be used as an estimate of $r(t)$. The above mentioned formula leads to an estimate of the survivor function:

$$G^*(t) = \prod_{\tau=0}^{t-1} (1 - r^*(\tau))$$

Basically the same procedure can be used if one assumes a continuous time axis. The resulting estimated survivor function is then a step function having steps at the points in time when events occur.

Exercise 4.3 Using the data in Box 4.1, calculate a survivor function $G^*(t)$ with the Kaplan-Meier procedure. Observe that the calculation of $r^*(t)$ is only required for points in time at which events occur.

4.4 Multiple Destination States

The Kaplan-Meier procedure can also be used if episodes can end in two or more destination states. The procedure leads to estimates of sub-survivor functions

$$G_d^*(t) = \prod_{\tau=0}^{t-1} (1 - r_d^*(\tau))$$

where $d \in \mathcal{D}$. They are also called *pseudo survivor functions* because they cannot be interpreted as survivor functions (in its normal sense). Interpretations should be based on state-specific rate functions which can be estimated by

$$r_d^*(t) = \frac{E_d^*(t)}{R^*(t)}$$

where $E_d^*(t)$ now denotes the number of episodes ending at t in destination state d.

Chapter 5

Regression Models for States

We now begin with a discussion of models that can be used to consider relationships between variables. As a starting point, remember the two forms for the representation of life course data: A sequence of state variables, Y_t, with t referring to a discrete time axis, and a two-dimensional duration variable (T, D). Both forms can be used for the development of models. This chapter assumes the first form.

5.1 The Modeling Approach

We assume a sequence of state variables

$$Y_t : \Omega \longrightarrow \mathcal{Y}$$

defined for a discrete process time axis $t = 0, 1, 2, \ldots$ The state space consists of two or more states. The state space will be denoted by $\mathcal{Y}$. When the process has developed until time t, its statistical representation is given by the distribution

$$\mathrm{P}(Y_t = y_t, Y_{t-1} = y_{t-1}, \ldots, Y_0 = y_0)$$

where $y_0, \ldots, y_t$ are the possible states in $\mathcal{Y}$.

Models serve to think about relationships between variables. In order to develop our modeling approach, the following abbreviation will be helpful:

$$\bar{Y}_t := (Y_t, Y_{t-1}, \ldots, Y_0)$$

It will be called a *process variable*. Possible values will be denoted by corresponding lower-case values

$$\bar{y}_t := (y_t, y_{t-1}, \ldots, y_0)$$

They represent possible sequences of states. The starting point for the model construction can then be denoted by

$$\mathrm{P}(\bar{Y}_t = \bar{y}_t)$$

In order to assess dependence relations between the state variables, one can use conditional distributions. The basic idea is that the process develops sequentially in time, beginning with a given initial distribution. This can be symbolically depicted as

$$\begin{aligned}
&Y_0 \\
&Y_1 \,|\, Y_0 \\
&Y_2 \,|\, Y_0, Y_1 \\
&\quad\vdots \\
&Y_t \,|\, Y_0, Y_1, \ldots, Y_{t-1}
\end{aligned}$$

Then, by successively creating conditional distributions, one gets

$$\mathrm{P}(\bar{Y}_t = \bar{y}_t) = \prod_{\tau=1}^{t} \mathrm{P}(Y_\tau = y_\tau \,|\, \bar{Y}_{\tau-1} = \bar{y}_{\tau-1})\, \mathrm{P}(Y_0 = y_0) \tag{5.1}$$

5.2 Theoretical Speculations

The general approach (5.1) can be used in two different ways as a starting point for further considerations. It can be used for speculations about possible forms of processes, and it can be used as a formal framework for the representation of given data. We begin with a short discussion of some theoretical possibilities.

Since we assume that the initial distribution of Y_0 is given, speculations concern the conditional distributions

$$\mathrm{P}(Y_\tau = y_\tau \,|\, \bar{Y}_{\tau-1} = \bar{y}_{\tau-1})$$

A particularly simple assumption is a one-step memory:

$$\mathrm{P}(Y_\tau = y_\tau \,|\, \bar{Y}_{\tau-1} = \bar{y}_{\tau-1}) \;=\; \mathrm{P}(Y_\tau = y_\tau \,|\, Y_{\tau-1} = y_{\tau-1})$$

It is assumed that the state in some temporal location t only depends on the state in the temporal location $t-1$. Somewhat more complicated would be a two-step memory

$$\begin{aligned}
&\mathrm{P}(Y_\tau = y_\tau \,|\, \bar{Y}_{\tau-1} = \bar{y}_{\tau-1}) \;= \\
&\qquad \mathrm{P}(Y_\tau = y_\tau \,|\, Y_{\tau-1} = y_{\tau-1}, Y_{\tau-2} = y_{\tau-2})
\end{aligned}$$

Exercise 5.1 Consider a process with a one-step memory. The state space is $\mathcal{Y} = \{0, 1\}$, initially all individuals are in state 0, and the transition probabilities are given by

$$\begin{aligned}
\mathrm{P}(Y_\tau = 1 \,|\, Y_{\tau-1} = 0) \;&=\; 1/2 \\
\mathrm{P}(Y_\tau = 1 \,|\, Y_{\tau-1} = 1) \;&=\; 1/3
\end{aligned}$$

For $t = 0, \ldots, 8$, using a die, construct 10 realizations of the process and present the development of the state distribution in form of a table.

Exercise 5.2 Assume that there are 5 distinct states. How many transition probabilities would be necessary in order to completely specify a process with a two-step memory?

5.3 Models with Two States

We consider processes with two possible states, $\mathcal{Y} = \{0, 1\}$, and furthermore assume a one-step memory. Without further restrictions, a model requires two parameters for each point in time:

$$\begin{aligned} \mathrm{P}(Y_t = 1 \,|\, Y_{t-1} = 0) &= \theta_{10,t} \\ \mathrm{P}(Y_t = 1 \,|\, Y_{t-1} = 1) &= \theta_{11,t} \end{aligned}$$

The quantities $\theta_{10,t}$ and $\theta_{11,t}$ are called *parameters of the process*. In general, these parameters can change while the process is going on. As a radical simplification, one can assume that they will not change. The process is then called *stationary*, and the modeling approach becomes

$$\begin{aligned} \mathrm{P}(Y_t = 1 \,|\, Y_{t-1} = 0) &= \theta_{10} \\ \mathrm{P}(Y_t = 1 \,|\, Y_{t-1} = 1) &= \theta_{11} \end{aligned}$$

Exercise 5.3 Using the data in Box 2.1, calculate the time-varying process parameters

$$\theta_{ij,t} \qquad \text{for} \quad i,j \in \{0,1\},\ t = 1,\ldots,5$$

Then assume that the data result from a stationary process and calculate the time-constant process parameters

$$\theta_{ij} \qquad \text{for} \quad i,j \in \{0,1\}$$

5.4 Models with Covariates

So far we considered a single state variable, Y_t, and models concerned the question how its distribution might depend on earlier values of the same state variable. Often, one is (also) interested in how the distribution of Y_t depends on values of other variables which are then called *covariates*. There are two kinds of such covariates:

- *time-constant covariates* having values that are fixed at the beginning of the process and do not change while the process is going on; and
- *time-varying covariates* having values that might change while the process is going on.

Box 5.1 Data set 5

ID	t =	0	1	2	3	4	5	6
1	Y	0	0	0	1	1	0	0
	X1	0	0	0	0	0	0	0
	X2	20	21	22	23	24	25	26
2	Y	1	1	0	0	0	1	1
	X1	0	0	0	0	0	0	0
	X2	22	23	24	25	26	27	28
3	Y	1	1	1	0	0	0	0
	X1	0	0	0	0	0	0	0
	X2	21	22	23	24	25	26	27
4	Y	0	0	0	0	1	1	1
	X1	1	1	1	1	1	1	1
	X2	20	21	22	23	24	25	26
5	Y	0	0	1	1	1	0	0
	X1	1	1	1	1	1	1	1
	X2	22	23	24	25	26	27	28
6	Y	1	1	1	0	0	1	1
	X1	1	1	1	1	1	1	1
	X2	21	22	23	24	25	26	27

Since time-constant covariates can be considered as a special case of time-varying covariates, we consider only the more general case. The idea of parallel processes provides a useful framework. The process of primary interest is given by Y_t. A covariate process defined on the same time axis is given by X_t, i.e:

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

As before, we assume that Y_t is a discrete one-dimensional state variable. X_t can be a multidimensional variable with an arbitrary property space, for example, an m-dimensional variable

$$X_t = (X_{t1}, \ldots, X_{tm})$$

However, to ease notations, we assume that also X_t is a one-dimensional discrete variable.

The idea is that the distribution of Y_t depends not only on previous state variables but also on previous values of covariates. For the discussion of models, we shall assume that covariates will not depend on state variables (covariates are then called *exogenous*). The modeling framework (5.1) can then be extended as follows:

$$\mathrm{P}(\bar{Y}_t = \bar{y}_t) = \prod_{\tau=1}^{t} \mathrm{P}(Y_\tau = y_\tau \,|\, \bar{Y}_{\tau-1} = \bar{y}_{\tau-1}, \bar{X}_{\tau-1} = \bar{x}_{\tau-1})\ \mathrm{P}(Y_0 = y_0, X_0 = x_0) \tag{5.2}$$

Again, this framework can be simplified in many different ways. For example, assuming a one-step memory also for the dependence on covariates, would lead to

$$\begin{aligned} \mathrm{P}(Y_\tau = y_\tau \,|\, \bar{Y}_{\tau-1} = \bar{y}_{\tau-1}, \bar{X}_{\tau-1} = \bar{x}_{\tau-1}) \;=\; & \\ \mathrm{P}(Y_\tau = y_\tau \,|\, Y_{\tau-1} = y_{\tau-1}, X_{\tau-1} = x_{\tau-1}) & \end{aligned}$$

As an example, we consider the data set 5 in Box 5.1. These data contain information about the development of states for six persons. There are two possible states, $\mathcal{Y} = \{0, 1\}$, and two covariates. The covariate X_1 is time-constant, e.g., the sex (0 = men, 1 = women); the covariate X_2 is time-varying, e.g., the age of persons.

5.5 Binary Logit Models

Modeling processes with covariates is often done with logit models. We here discuss these models for processes with only two states, $\mathcal{Y} = \{0, 1\}$. The general model with time-varying parameters is

$$\mathrm{P}(Y_t = 1 \,|\, Y_{t-1} = 0, X_{t-1} = x_{t-1}) \;=\; \frac{\exp(\alpha_{10,t} + x_{t-1}\beta_{10,t})}{1 + \exp(\alpha_{10,t} + x_{t-1}\beta_{10,t})}$$

$$\mathrm{P}(Y_t = 1 \,|\, Y_{t-1} = 1, X_{t-1} = x_{t-1}) \;=\; \frac{\exp(\alpha_{11,t} + x_{t-1}\beta_{11,t})}{1 + \exp(\alpha_{11,t} + x_{t-1}\beta_{11,t})}$$

Assuming a stationary process, one can drop the references to time on the right-hand side. Obviously, the model without covariates discussed in Section 5.3 is a special case.

Exercise 5.4 Draw the graph of the logit function

$$z = \frac{\exp(x)}{1 + \exp(x)}$$

in the range $-3 \le x \le 3$.

Exercise 5.5 Using the notation of exercise 5.4, derive the inverse function that shows how x depends on z. Calculate the values of x that correspond to $z = 0.5, 0.6,$ and 0.7.

Exercise 5.6 Write down a logit model for the data in Box 5.1 that assumes a stationary process. Indicate the model parameters that should be estimated.

5.6 Maximum Likelihood Estimation

We briefly discuss the maximum likelihood estimation of the parameters of a logit model. Assuming a model for a stationary process, one has to estimate four parameters:

$$\alpha_{10}, \beta_{10}, \alpha_{11}, \beta_{11}$$

It will be assumed that data are given as

$$(x_{it}, y_{it}) \qquad \text{for} \quad i = 1, \ldots, N, \; t = 0, \ldots, T$$

The model then implies

$$\mathrm{P}(Y_t = y_{i,t} \,|\, Y_{t-1} = y_{i,t-1}, X_{t-1} = x_{i,t-1}) =$$

$$\begin{cases} \frac{\exp(\alpha_{10} + x_{i,t-1}\beta_{10})}{1 + \exp(\alpha_{10} + x_{i,t-1}\beta_{10})} & \text{if } \; y_{i,t} = 1, y_{i,t-1} = 0 \\ \frac{1}{1 + \exp(\alpha_{10} + x_{i,t-1}\beta_{10})} & \text{if } \; y_{i,t} = 0, y_{i,t-1} = 0 \\ \frac{\exp(\alpha_{11} + x_{i,t-1}\beta_{11})}{1 + \exp(\alpha_{11} + x_{i,t-1}\beta_{11})} & \text{if } \; y_{i,t} = 1, y_{i,t-1} = 1 \\ \frac{1}{1 + \exp(\alpha_{11} + x_{i,t-1}\beta_{11})} & \text{if } \; y_{i,t} = 0, y_{i,t-1} = 1 \end{cases}$$

In a more compact notation:

$$\mathrm{P}(Y_t = y_{i,t} \,|\, Y_{t-1} = y_{i,t-1}, X_{t-1} = x_{i,t-1}) =$$

$$\left(\frac{\exp(\alpha_{10} + x_{i,t-1}\beta_{10})^{y_{it}}}{1 + \exp(\alpha_{10} + x_{i,t-1}\beta_{10})} \right)^{1 - y_{i,t-1}} \left(\frac{\exp(\alpha_{11} + x_{i,t-1}\beta_{11})^{y_{it}}}{1 + \exp(\alpha_{11} + x_{i,t-1}\beta_{11})} \right)^{y_{i,t-1}}$$

Viewed as a function of the model parameters, this is called the *likelihood* of the ith observation at t. Combining these likelihoods for all observations, one gets the *likelihood function*

$$\begin{aligned} \mathcal{L}(\alpha_{10}, \beta_{10}, \alpha_{11}, \beta_{11}) \;=\; \prod_{i=1}^{N} \prod_{t=1}^{T} & \left(\frac{\exp(\alpha_{10} + x_{i,t-1}\beta_{10})^{y_{it}}}{1 + \exp(\alpha_{10} + x_{i,t-1}\beta_{10})} \right)^{1 - y_{i,t-1}} \\ & \left(\frac{\exp(\alpha_{11} + x_{i,t-1}\beta_{11})^{y_{it}}}{1 + \exp(\alpha_{11} + x_{i,t-1}\beta_{11})} \right)^{y_{i,t-1}} \end{aligned}$$

Maximizing this function, one gets the ML estimates of the model parameters, denoted by

$$\hat{\alpha}_{10}, \hat{\beta}_{10}, \hat{\alpha}_{11}, \hat{\beta}_{11}$$

Exercise 5.7 Calculate the log-likelihood function

$$\ell(\alpha_{10}, \beta_{10}, \alpha_{11}, \beta_{11}) \;=\; \log\left(\mathcal{L}(\alpha_{10}, \beta_{10}, \alpha_{11}, \beta_{11})\right)$$

Exercise 5.8 Calculate the value of the log-likelihood function derived in exercise 5.7 for the data in Box 5.1 and for the parameter values

$$\alpha_{10} = \beta_{10} = \alpha_{11} = \beta_{11} = 0$$

Box 5.2 Data set 5a

ID	t	Y(t)	Y(t-1)	X1(t-1)	X2(t-1)
1	1	0	0	0	20
1	2	0	0	0	21
1	3	1	0	0	22
1	4	1	1	0	23
1	5	0	1	0	24
1	6	0	0	0	25
2	1	1	1	0	22
2	2	0	1	0	23
2	3	0	0	0	24
2	4	0	0	0	25
2	5	1	0	0	26
2	6	1	1	0	27
3	1	1	1	1	21
3	2	1	1	1	22
3	3	1	0	1	23
3	4	0	0	1	24
3	5	0	0	1	25
3	6	0	0	1	26
4	1	0	0	1	20
4	2	0	0	1	21
4	3	0	0	1	22
4	4	1	0	1	23
4	5	1	1	1	24
4	6	1	1	1	25
5	1	0	0	1	22
5	2	1	0	1	23
5	3	1	1	1	24
5	4	1	1	1	25
5	5	0	1	1	26
5	6	0	0	1	27
6	1	1	1	1	21
6	2	1	1	1	22
6	3	0	1	1	23
6	4	0	0	1	24
6	5	1	0	1	25
6	6	1	1	1	26

Exercise 5.9 ML estimates of model parameters result from the maximization of the log-likelihood function. In order to find the values, one can use standard packages that allow to estimate simple logit models. This is based on the fact that the likelihood function can be partitioned into two factors that can be maximized separately. Describe how to reorganize the data set (see Box 5.2).

Exercise 5.10 Using for X the age variable (`X2` in Box 5.1), one gets the following ML estimates of the model parameters:

$$\hat{\alpha}_{10} = -3.14,\ \hat{\beta}_{10} = 0.10,\ \hat{\alpha}_{11} = 4.91,\ \hat{\beta}_{11} = -0.16$$

A computer program (e.g. TDA) computes the maximized log-likelihood function as -12.14 (for the first half model, $Y_{t-1} = 0$) and -8.88 (for the second half model, $Y_{t-1} = 1$). Calculate the values of the log-likelihood function of the combined model.

Exercise 5.11 Using the parameter estimates mentioned in exercise 5.10, create a table that shows for age values $X = 20, \ldots, 26$ and values $Y_{t-1} = 0, 1$ the estimated probabilities

$$\mathrm{P}(Y_t = 0 \,|\, Y_{t-1} = \cdots, X = \cdots) \quad \text{and}$$
$$\mathrm{P}(Y_t = 1 \,|\, Y_{t-1} = \cdots, X = \cdots)$$

Chapter 6

Models for Durations

We consider single episodes, represented by a two-dimensional statistical variable

$$(T, D) \qquad \text{with } T \in \mathcal{T},\ D \in \mathcal{D}$$

$\mathcal{D}$ is the set of possible destination states, $\mathcal{T}$ is the (discrete or continuous) time axis. For the following discussion, we assume a continuous time axis and identify $\mathcal{T}$ with the set of nonnegative real numbers.

Statistical models can serve, both, to represent data and to formulate (more general) hypotheses that concern how episodes could, or probably will, develop. One can distinguish, respectively, between descriptive and analytical models.

6.1 Time-constant Rates

For episodes with a single destination state it suffices to consider a duration variable T. A statistical model consists in making an assumption about the variable's distribution. In order to characterize the distribution, one can use, equivalently, a distribution function $F(t)$, a survivor function $G(t)$, a density function $f(t)$, or a rate function $r(t)$.

In order to formulate assumptions, it is often easiest to use the rate function. The simplest assumption is to assume a constant rate:

$$r(t) = \theta$$

where θ is a parameter that can vary in some specified set (parameter space), in the following denoted by Θ. Since rates are nonnegative, we identify Θ with the set of nonnegative real numbers.

A model that is based on the assumption of a time-constant rate is called an *exponential model*. The distribution is called an *exponential distribution* with parameter θ.

Using the formula for the relationship between rate and survivor function, the survivor function of the exponential distribution is given by

$$G(t) = \exp\left\{-\int_0^t r(\tau)\, d\tau\right\} = \exp(-\theta t)$$

Exercise 6.1 Derive formulas for the distribution function and for the density function of the exponential distribution with parameter θ.

Exercise 6.2 Draw graphs of the distribution function and of the density function of the standard exponential distribution ($\theta = 1$).

6.2 Weibull Distribution

Most models use rates that change in some specific way with time. Often used is the *Weibull distribution* which implies a rate function that can monotonically rise or decrease. The Weibull distribution has two parameters; the survivor function is

$$G(t) = \exp\{-(\alpha t)^\beta\}$$

It is assumed that both parameters, α and β, can only take positive values. Differentiating the survivor function, one gets the density function

$$f(t) = \beta\alpha^\beta t^{\beta-1} \exp\{-(\alpha t)^\beta\}$$

and the rate function

$$r(t) = \beta\alpha^\beta t^{\beta-1}$$

Exercise 6.3 Show step by step how one can derive the density function and the rate function from the survivor function of the Weibull distribution.

Exercise 6.4 For which parameter values does one get the exponential distribution as a special case of the Weibull distribution?

Exercise 6.5 Based on a time axis from 0 to 3, draw the graph of the Weibull rate function for parameter values $\alpha = 1$ and $\beta = 0.5, 1.0, 1.5$.

6.3 Log-Logistic Distribution

A simple model with non-monotonic rate functions is based on the log-logistic distribution. The distribution has two parameters, α and β, restricted to positive values. The survivor function is

$$G(t) = \frac{1}{1 + (\alpha t)^\beta}$$

Differentiation leads to the density function

$$f(t) = \frac{\beta\alpha^\beta t^{\beta-1}}{(1 + (\alpha t)^\beta)^2}$$

and one then finds the the rate function

$$r(t) = \frac{\beta\alpha^\beta t^{\beta-1}}{1 + (\alpha t)^\beta}$$

Exercise 6.6 Show step by step how one can derive the density function and the rate function from the survivor function of the log-logistic distribution.

Exercise 6.7 Show that the exponential distribution is *not* a special case of the log-logistic distribution.

Exercise 6.8 Based on a time axis from 0 to 3, draw the graph of the log-logistic rate function for parameter values $\alpha = 1$ and $\beta = 1$ and 2.

Exercise 6.9 Show that the log-logistic rate function (if it is concave) has its maximum at

$$t_{\max} = \frac{1}{\alpha}(\beta - 1)^{\frac{1}{\beta}}$$

For which parameter values is the rate function concave? Which is the value of the maximum?

6.4 Log-Normal Distribution

Standard regression models often assume a normal distribution (of residuals). Since the duration variable T can only take positive values, one might assume that its logarithm is normally distributed.[1] It is then said that T has a *log-normal distribution*.

In order to follow this idea, we begin with a general consideration. Let X denote a continuous variable with distribution function $F_X(x)$ and density function $f_X(x)$. Also, let g denote an arbitrary monotonically increasing function. One can then consider the variable

$$Y = g(X)$$

and ask how to derive the distribution and density functions of Y, i.e., respectively, $F_Y(y)$ and $f_Y(y)$, from the corresponding functions of X.[2] For the distribution functions, one can use the relationship

$$F_X(x) = \mathrm{P}(X \leq x) = \mathrm{P}(Y \leq g(x)) = F_Y(g(x))$$

For the density functions, one finds

$$f_X(x) = \left.\frac{dF_X(u)}{du}\right|_{u=x} = \left.\frac{dF_Y(g(u))}{du}\right|_{u=x} = f_Y(g(x))g'(x)$$

with $g'(x)$ denoting the derivative of the transformation function g.

[1] We always mean the natural logarithm, i.e., the inverse function of the exponential.

[2] Since g is monotonic, also the inverse relationship

$$X = g^{-1}(Y)$$

exists; the problem if obviously symmetrical.

Now assume that Y is normally distributed, and the relationship with the duration variable T is given by

$$Y = \log(T)$$

The density function of Y is then

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right\}$$

Applying the transformation derived above, one gets the density function of T, now denoted by $f(t)$, as follows:

$$f(t) = \frac{1}{\sqrt{2\pi}\sigma t} \exp\left\{-\frac{1}{2}\left(\frac{\log(t)-\mu}{\sigma}\right)^2\right\}$$

This is the density function of the log-normal distribution with parameters μ and σ. μ can take any, σ only positive values.

Exercise 6.10 Based on a time axis from 0 to 3, draw the graph of the log-normal rate function for parameter values $\mu = 0$ and $\sigma = 1$.

6.5 Multiple Destination States

In order to represent episodes with multiple destination states, one needs a two-dimensional variable (T, D). Its distribution can be characterized by state-specific rates

$$r_d(t) \qquad \text{for } d \in \mathcal{D}$$

Models can be specified by simply using the approaches discussed in the foregoing sections separately for each state-specific rate function. For example, an exponential model would assume

$$r_d(t) = \theta_d$$

6.6 Mixture Distributions

Mixture models result from the assumption that the collection of individuals, Ω, consists of two or more parts with different rate functions. Assume that there are m parts and that, at $t = 0$, the proportion of the jth part is given by π_j, implying that

$$\sum_{j=1}^{m} \pi_j \ = \ 1$$

For each part j, $r_j(t)$ denotes the rate function, $f_j(t)$ the density function, and $G_j(t)$ the survivor function. The temporal development of the proportion of the jth part is then given by

$$\pi_j G_j(t) \Big/ \sum_{k=1}^{m} \pi_k G_k(t)$$

where

$$G(t) = \sum_{k=1}^{m} \pi_k G_k(t)$$

is the overall survivor function. This allows to derive the density function

$$f(t) = -\frac{dG(t)}{dt} = \sum_{j=1}^{m} \pi_j \left(-\frac{dG_j(t)}{dt} \right) = \sum_{j=1}^{m} \pi_j f_j(t)$$

Finally, one finds the mean rate in Ω as

$$r(t) = \frac{\sum_{j=1}^{m} \pi_j f_j(t)}{\sum_{j=1}^{m} \pi_j G_j(t)}$$

Exercise 6.11 Assume two parts and $\pi_1 = \pi_2 = 0.5$. Assume also time-constant rates $r_1(t) = 1$ and $r_2(t) = 2$, respectively. Derive the development of the mean rate and show that it decreases with time.

Chapter 7

Rate Models with Covariates

This chapter considers rate models with time-constant covariates. Formulations are now based on a statistical variable (T, D, X) representing single episodes. T denotes durations, D denotes the destination state (or is zero if an observation is right censored), and X (possibly multidimensional) represents the values of a time-constant covariate.

We begin with models for episodes having only a single destination state. Models for episodes with multiple destination states will be discussed in Section 7.4.

7.1 The Exponential Model

The exponential model assumes a time-constant rate

$$r(t) = \theta$$

The basic idea is to make this rate dependent on values of covariates. Normally, one uses a link function that guarantees that the rate will be positive. The standard model uses an exponential link function:

$$r(t \,|\, X = x) \;=\; \exp(\beta_0 + x\beta_1)$$

If there are m covariates $(X_1, \ldots, X_m)$, a general formulation is

$$r(t \,|\, X_1 = x_1, \ldots, X_m = x_m) \;=\; \exp(\beta_0 + x_1\beta_1 + \ldots + x_m\beta_m)$$

Exercise 7.1 Derive the following formula for the mean of an exponential distribution with parameter θ:[1]

$$E(T) = \int_0^\infty t f(t)\, dt = \int_0^\infty \theta t \exp(-\theta t)\, dt = \frac{1}{\theta}$$

Exercise 7.2 Find a formula for the median of an exponential distribution with parameter θ.

[1] Use the following rule for partial integration:

$$\int F(t)g(t)\, dt = F(t)G(t) - \int f(t)G(t)\, dt$$

where $f(t) = dF(t)/dt$ and $g(t) = dG(t)/dt$. Use $F(t) = t$, $g(t) = \theta \exp(-\theta t)$.

Box 7.1 Data set 6

```
ID    T   D   X1   X2
----------------------
 1   17   1    8    1
 2    5   0    5    0
 3   22   1    9    1
 4   13   1    7    0
 5    2   0    5    0
 6    9   1    6    1
 7   12   0    5    1
 8   15   1    7    1
```

7.2 Estimation of Model Parameters

Assume that n observations

$$(t_i, d_i, x_{i1}, \ldots, x_{im}) \qquad \text{for } i = 1, \ldots, n$$

are given. t_i is the value of the duration variable T, d_i shows whether the observation is right censored ($d_i = 0$) or not ($d_i = 1$), and $x_{i1}, \ldots, x_{im}$ are the values of the covariates.

Models are estimated with the maximum likelihood approach. Using the density

$$f(t_i \mid X_1 = x_{i1}, \ldots, X_m = x_{im})$$

for uncensored observations and the survivor function

$$G(t_i \mid X_1 = x_{i1}, \ldots, X_m = x_{im})$$

for censored observations, the likelihood function looks as follows:

$$\mathcal{L}(\beta_0, \ldots, \beta_m) = \prod_{i=1}^{n} f(t_i \mid X_1 = x_{i1}, \ldots, X_m = x_{im})^{d_i} \\ G(t_i \mid X_1 = x_{i1}, \ldots, X_m = x_{im})^{1-d_i}$$

Since $r(t) = f(t)/G(t)$, the likelihood function can also be written in the following form:

$$\mathcal{L}(\beta_0, \ldots, \beta_m) = \prod_{i=1}^{n} r(t_i \mid X_1 = x_{i1}, \ldots, X_m = x_{im})^{d_i} \\ G(t_i \mid X_1 = x_{i1}, \ldots, X_m = x_{im})$$

We now consider the exponential model introduced in Section 7.1. One then

gets

$$r(t_i \mid X_1 = x_{i1}, \ldots, X_m = x_{im}) = \exp(\beta_0 + x_{i1}\beta_1 + \ldots + x_{im}\beta_m)$$

$$G(t_i \mid X_1 = x_{i1}, \ldots, X_m = x_{im}) = \\ \exp\{-\exp(\beta_0 + x_{i1}\beta_1 + \ldots + x_{im}\beta_m)\, t_i\}$$

and finds the likelihood function

$$\mathcal{L}(\beta_0, \ldots, \beta_m) = \prod_{i=1}^{n} \exp(\beta_0 + x_{i1}\beta_1 + \ldots + x_{im}\beta_m)^{d_i} \\ \exp\{-\exp(\beta_0 + x_{i1}\beta_1 + \ldots + x_{im}\beta_m)\, t_i\}$$

Maximization most often uses the corresponding log-likelihood function

$$\ell(\beta_0, \ldots, \beta_m) = \sum_{i=1}^{n} d_i(\beta_0 + x_{i1}\beta_1 + \ldots + x_{im}\beta_m) - \\ \exp(\beta_0 + x_{i1}\beta_1 + \ldots + x_{im}\beta_m)\, t_i$$

If the model contains covariates, it is normally not possible to find the maximum with analytical methods. Instead, one must use an iterative maximization procedure (that requires a computer).

A simple approach is possible, however, if the model does not contain covariates. Then, the log-likelihood function is

$$\ell(\beta_0) = \sum_{i=1}^{n} d_i\beta_0 - t_i \exp(\beta_0)$$

and one can derive the gradient (first derivative)

$$\frac{\partial \ell(\beta_0)}{\partial \beta_0} = \sum_{i=1}^{n} d_i - t_i \exp(\beta_0)$$

Finally, equating the gradient with zero leads to the ML estimate of β_0, namely

$$\hat{\beta}_0 = \log\left(\frac{\sum_{i=1}^{n} d_i}{\sum_{i=1}^{n} t_i}\right)$$

Exercise 7.3 Using the second derivative, show that the log-likelihood function of the exponential model without covariates has exactly one maximum.

Exercise 7.4 Based on the data in Box 7.1, estimate the time-constant rate of an exponential model without covariates. Derive an estimate of the mean duration.

7.3 A General Modeling Approach

Rate models can be based on many different rate functions. The exponential model with a time-constant rate is just one special case. The general formulation of a parametric rate model starts from a rate function that depends on a one- or multidimensional parameter θ. Rate, density, and survivor function can then be written as follows:

$$r(t \mid \theta),\ f(t \mid \theta),\ G(t \mid \theta)$$

In order to include covariates, one uses a link function that makes the distribution parameter θ dependent on values of the covariates. For example, if the distribution has two parameter, say $\theta = (\alpha, \beta)$, one can use link function having the following form:

$$\begin{aligned} \alpha &= g_\alpha(\alpha_0 + x_1\alpha_1 + \ldots + x_m\alpha_m) \\ \beta &= g_\beta(\beta_0 + x_1\beta_1 + \ldots + x_m\beta_m) \end{aligned}$$

Exercise 7.5 Derive the log-likelihood function for a Weibull model without covariates.

Exercise 7.6 Derive the log-likelihood function for a log-logistic model without covariates.

7.4 Multiple Destination States

The modeling approach introduced in the previous section can be generalized for episodes with multiple destination states. Starting point is a general rate function

$$r_d(t \mid \theta_d) \qquad \text{for } d \in \mathcal{D}$$

where d refers to possible destination states. From this, one gets the overall rate function

$$r(t \mid \theta) \ = \sum_{d \in \mathcal{D}} r_d(t \mid \theta_d)$$

with θ denoting the collection of destination-specific parameters θ_d. From the overall rate the formula of the survivor function becomes

$$G(t \mid \theta) \ = \ \exp\left\{ - \int_0^t r(\tau \mid \theta)\, d\tau \right\}$$

An equivalent formulation would be

$$G(t \mid \theta) \ = \prod_{d \in \mathcal{D}} G_d(t \mid \theta_d)$$

with

$$G_d(t \mid \theta_d) \ = \ \exp\left\{ - \int_0^t r_d(\tau \mid \theta_d)\, d\tau \right\}$$

Finally, in order to derive the likelihood function, we assume that the data are given as in Section 7.2 (with d_i now denoting the individual's destination state). Using the indicator function $I(d = d_i)$ (taking the value 1 if $d = d_i$ and zero otherwise), the likelihood function can be written in the following form:

$$\mathcal{L}(\theta) = \prod_{i=1}^{n} G(t_i \mid \theta) \prod_{d \in \mathcal{D}} r(t_i \mid \theta_d)^{I(d = d_i)}$$

Exercise 7.7 Derive the likelihood and the log-likelihood function of an exponential model with three destination states.

7.5 Pseudo Residuals

In contrast to simple regression models with a quantitative dependent variable, there is no simple method to judge the goodness-of-fit of a rate model. As an aid, one sometimes uses so-called pseudo residuals. In order to explain the idea, we refer to episodes with a single destination state. Data are assumed to be given as

$$(t_i, d_i, x_i) \qquad \text{for } i = 1, \ldots, n$$

We also assume that one has estimated a rate model so that the functions

$$r(t \mid x; \hat{\theta}),\ f(t \mid x; \hat{\theta}),\ G(t \mid x; \hat{\theta})$$

can be computed, where $\hat{\theta}$ denotes the estimated model parameters.

This model can now be viewed as the description of a two-step random generator

$$\boxed{X} \xrightarrow{\ x\ } \boxed{T_x} \longrightarrow (t, x)$$

The first random generator generates a value x of the covariate vector according to the statistical distribution of X (given by the data). The second random generator generates a duration t, based on the model $f(t \mid x; \hat{\theta})$, i.e., conditional on the value x that was realized in the first step.

This method can be used to generate an arbitrary number of pseudo observations (t_j^*, x_j^*) $(j = 1, 2, 3, \ldots)$, such that the distribution of the values x_j^* equals the distribution of X, and the conditional distribution of the durations t_j^* conforms with the model. Accordingly, for each possible value x in the property space of X, one can consider the generated durations as realizations of a random variable T_x having a distribution defined by $f(t \,|\, x; \hat{\theta})$.

In a second step, one considers a transformation of the random variable T_x such that it no longer depends on specific values of the covariate vector. Using the transformation

$$T_x \longrightarrow J(T_x)\,, \text{ defined by } t \longrightarrow J(t) = \int_0^t r(\tau \mid x; \hat{\theta})\, \mathrm{d}\tau$$

results in a standard exponential distribution of $J(T_x)$ (having the constant rate 1). This can be seen in the following way:

$$\begin{aligned} \mathrm{P}\left(J(T_x) > t\right) &= \mathrm{P}\left(T_x > J^{-1}(t)\right) \\ &\equiv G(J^{-1}(t) \mid x; \hat{\theta}) \\ &= \exp\Big\{ - \int_0^{J^{-1}(t)} r(\tau \mid x; \hat{\theta})\, \mathrm{d}\tau \Big\} \\ &= \exp\left\{ -J(J^{-1}(t)) \right\} \\ &= \exp(-t) \end{aligned}$$

The survivor function of the transformed random variable $J(T_x)$ equals the survivor function of a standard exponential distribution and consequently is independent of x and θ.[2]

This then allows to check the hypothesis that the data can be viewed as a random sample from the random generator described by the estimated model. If the hypothesis is true, the procedure should lead to a set of values

$$e_i = J(t_i)$$

having approximately a standard exponential distribution. The values e_i are called *pseudo residuals* (sometimes also *generalized residuals*). In order to check whether they follow a standard exponential distribution, one can calculate, and then check, their survivor function (of course, taking into account right censored observations). If $G_r(t)$ denotes the survivor function, one can consider the graph

$$t \longrightarrow -\log\left\{G_r(t)\right\}$$

If the residuals follow a standard exponential distribution, this graph should be approximately equal a 45° line.

Exercise 7.9 ________________

[2]Note that the distribution of T_x is defined through $G(\cdot \,|\, x, \theta)$, not through $G(\cdot \,|\, x, \hat{\theta})$. The procedure is, however, based on the hypothesis that θ has been correctly estimated.

Exercise 7.8 Derive a formula for the calculation of pseudo residuals for an exponential model with covariates.

Calculate pseudo residuals for the exponential model without covariates that was estimated in exercise 7.4.

Chapter 8

Time-varying Covariates

This chapter considers time-varying covariates defined as covariates that may change their values during an episode.

8.1 Conditional Survivor Functions

Remember the relationship between the rate and the survivor function:

$$G(t) = \exp\left\{-\int_0^t r(\tau)\,d\tau\right\}$$

Accordingly, we define a *conditional survivor function* by

$$G(t\,|\,s) = \exp\left\{-\int_s^t r(\tau)\,d\tau\right\}$$

It follows that

$$G(t) = G(t\,|\,s)\,G(s)$$

Such a partition can be repeated. Assume a division into k subintervals:

$$0 = t_0 < t_1 < \cdots < t_{k-1} < t_k = t$$

One then finds:

$$G(t) = \prod_{j=1}^{k} G(t_j\,|\,t_{j-1})$$

8.2 Reformulation of the Likelihood Function

Remember the likelihood function for the estimation of a rate model for a single episode with one destination state:

$$\mathcal{L}(\theta) = \prod_{i=1}^{n} r(t_i\,|\,\theta)^{d_i}\,G(t_i\,|\,\theta)$$

For each individual i, the duration t_i can be divided into an arbitrary number of subintervals:

$$0 = t_{i,0} < t_{i,1} < \cdots < t_{i,k_i-1} < t_{i,k_i} = t_i$$

The likelihood function can then be written as follows:

$$\mathcal{L}(\theta) = \prod_{i=1}^{n} r(t_i\,|\,\theta)^{d_i} \prod_{j=1}^{k_i} G(t_{i,j}\,|\,t_{i,j-1},\theta)$$

This will be called a likelihood function based on episode splitting.

Exercise 8.1 Starting from an arbitrary division of the process time axis into subintervals, write down a likelihood function for the estimation of a simple exponential model, first with, then without episode splitting. Show that both lead to identical parameter estimates.

8.3 Time-varying 0-1-Variables

We now assume that the data contain time-varying covariates and are given in the following form:

$$(t_i, d_i, x_i, z_{i,1}(t), \ldots, z_{i,m}(t)) \qquad \text{for } \ i = 1, \ldots, n$$

x_i is a vector of time-constant covariates, the variables $z_{i,j}(t)$ are time-varying. It will be assumed that these are 0-1-variables. $t_{i,j}$ denotes the point in time when $z_{i,j}(t)$ changes its value from 0 to 1.

Now consider all points in time when at least one covariate changes its value and assume that these points in time are ordered as follows:

$$0 = \tau_{i,0} < \tau_{i,1} < \cdots < \tau_{i,k_i-1} < \tau_{i,k_i} = t_i$$

The likelihood function can then be written in the following form:

$$\mathcal{L}(\theta) = \prod_{i=1}^{n} r(t_i\,|\,x_i, z_{i,1}(t_i), \ldots, z_{i,m}(t_i), \theta)^{d_i}$$

$$\prod_{j=1}^{k_i} G(\tau_{i,j}\,|\,\tau_{i,j-1}, x_i, z_{i1}(\tau_{i,j-1}), \ldots, z_{i,m}(\tau_{i,j-1}), \theta)$$

Box 8.1 Data set 7: Episode Splitting

```
ID  ORG  DES  TS  TF        ID  SPN  ORG  DES  TS  TF
--------------------        -------------------------
 1    0    1   0  10         1    1    0    0   0   3
 2    0    0   0  12         1    2    0    0   3   6
                             1    3    0    1   6  10
                             2    1    0    0   0   5
                             2    2    0    0   5   9
                             2    3    0    0   9  11
                             2    4    0    0  11  12
```

Box 8.2 Data set 8: Episode Splitting

```
ID  ORG  DES  TS  TF  Z     ID  SPN  ORG  DES  TS  TF   D
-----------------------     -----------------------------
 1    0    1   0  10  7      1    1    0    0   0   7   0
 2    0    1   0   8 -1      1    2    0    1   7  10   1
 3    0    1   0   5  8      2    1    0    1   0   0   1
 4    0    0   0  12  9      3    1    0    1   0   5   0
                             4    1    0    0   0   9   0
                             4    2    0    0   9  12   1
```

8.4 Episode Splitting

The approach just depicted is called the method of episode splitting (for the incorporation of time-varying variables into rate models). Box 8.1 illustrates the method with two episodes without covariates. The episodes are arbitrarily split respectively into three and four parts.

Box 8.2 illustrates the method with a time-varying covariate called `Z`. This variable indicates the point in time when the corresponding 0-1-variable (`D`) changes its value from 0 to 1. The right half of the box shows the splitted data set.

References

Blossfeld, H.-P., Rohwer, G. 2002. Techniques of Event History Modeling. New Approaches to Causal Analysis. 2nd Ed. Mahwah. NJ: Lawrence Erlbaum.

Cox, D. R., Oakes, D. 1984. Analysis of Survival Data. London: Chapman and Hall.

International Statistical Institute 1986. Declaration on Professional Ethics. International Statistical Review 54, 227–242.

Kaplan, E. L., Meier, P. 1958. Nonparametric Estimation from Incomplete Observations. Journal of the American Statistical Association 53, 457–481.

Lawless, J. F. 1982. Statistical Models and Methods for Lifetime Data. New York: Wiley.