Chapter 3

Parametric Survival Model with an Interval-Censored Covariate

This chapter is about the accelerated failure time model with an interval-censored covariate. Two cases concerning the response variable Y are distinguished: On one hand, Y is possibly right- or left-censored, on the other hand, Y is a doubly-censored variable.

In Section 3.1, we recall some of the properties of this parametric survival model, before presenting the particular model with an interval-censored covariate together with the observable data in Section 3.2. The two subsequent sections deal with the resulting likelihood functions (Section 3.3) and two particular parametric choices concerning the response variable's distribution (Section 3.4). Our proposal for the maximization of the likelihood functions by means of optimization methods is presented in Section 3.5. Finally, in Section 3.6, further aspects such as the inclusion of other covariates in the model and the consideration of data with missing values are addressed.

3.1 General aspects of the accelerated failure time model

Many models have been developed to detect possible predictors of a survival time. The probably most well-known one is Cox's semi-parametric proportional hazards model (1.5), which does not specify the survival time's distribution. In contrast with that model, the accelerated failure time model requires a specific distribution of the survival time, but does not assume proportional hazards. This model can be expressed in the following way modeling the survival function of time Y of a subject with covariate vector $\boldsymbol{Z}$:

$$S(y|\boldsymbol{Z}) = S_0\big(y \exp(\boldsymbol{\vartheta}'\boldsymbol{Z})\big), \tag{3.1}$$

where S_0 is the baseline survival function and $\boldsymbol{\vartheta}$ is the unknown parameter vector quantifying the influence of the covariates on Y. The term $\mathrm{AF} = \exp(-\boldsymbol{\vartheta}'\boldsymbol{Z})$ is called the acceleration factor, which relates the percentiles of an individual with covariate vector $\boldsymbol{Z}$, $y_p(\boldsymbol{Z})$, and the baseline percentiles $y_p(0)$ as follows:

$$y_p(\boldsymbol{Z}) = y_p(0)\exp(-\boldsymbol{\vartheta}'\boldsymbol{Z}). \tag{3.2}$$

This is deduced from the following equation:

$$p = 1 - S\big(y_p(\boldsymbol{Z})|\boldsymbol{Z}\big) = 1 - S_0\big(y_p(\boldsymbol{Z})\exp(\boldsymbol{\vartheta}'\boldsymbol{Z})\big) = 1 - S_0\big(y_p(0)\big).$$

Note that the above definition of the acceleration factor differs from the one in Klein and Moeschberger (1997, Chap. 12), in that they use the term $\exp(\boldsymbol{\vartheta}'\boldsymbol{Z})$ and not $\exp(-\boldsymbol{\vartheta}'\boldsymbol{Z})$.

According to equation (3.2), if a parameter ϑ is positive, the corresponding percentile $y_p(Z)$ is smaller than the baseline percentile. Consequently, on average, increasing Z implies deterioration of the survival time. On the other hand, if $\vartheta < 0$, Z is a protecting factor of the survival time. As an example, consider a dichotomic variable Z with parameter value $\vartheta = 0.7$. Then, the median survival time in group 1 is $\exp(-0.7) \approx 0.5$ times the median in group 0. On the other hand, if $\vartheta = -0.7$, the median of group 1 is twice the median of group 0.

As shown in Klein and Moeschberger, and in Gómez (2004, Tema 5), the accelerated failure time model can also be expressed in terms of a log linear model:

$$\ln(Y) = \mu + \boldsymbol{\beta}'\boldsymbol{Z} + \sigma W, \tag{3.3}$$

where $\boldsymbol{\beta}$ is the unknown parameter vector, σ the scale parameter, and W is the error term distribution. For example, if Y follows a Weibull distribution, W is the extreme value or Gumbel distribution; if Y follows a log logistic distribution, W is the standard logistic distribution.

Contrary to model (3.1), with the log linear expression, a positive parameter implies that the corresponding covariate is a protecting factor of the survival time. Both models, (3.1) and (3.3), are equivalent if S_0 is the survival function of the random variable $\exp(\mu + \sigma W)$ and $\boldsymbol{\beta} = -\boldsymbol{\vartheta}$. Hence, the term $\text{AF} = \exp(\boldsymbol{\beta}'\boldsymbol{Z})$ is the acceleration factor comparing an individual with covariate vector $\boldsymbol{Z}$ with a "baseline" individual.

The interpretation of the expression $\exp(-\boldsymbol{\beta}'\boldsymbol{Z}/\sigma)$ depends on the distribution of Y. In case of a Weibull distribution, this term is the relative risk (RR) of dying (being death the event of interest) of an individual with covariate vector $\boldsymbol{Z}$ compared to an individual with $\boldsymbol{Z} = \boldsymbol{0}$. The Weibull distribution is the only parametric choice for which the accelerated failure time model and the proportional hazards model are equivalent. If Y follows a log logistic distribution, $\text{OR} = \exp(-\boldsymbol{\beta}'\boldsymbol{Z}/\sigma)$ is the relative odds or odds ratio of dying comparing an individual with covariate vector $\boldsymbol{Z}$ and one with the baseline characteristics. The log logistic model is the only case, for which model (3.1) and the proportional odds model are equivalent. The latter model has the following expression:

$$\frac{S(y|\boldsymbol{Z})}{1 - S(y|\boldsymbol{Z})} = \exp(-\boldsymbol{\beta}'\boldsymbol{Z}/\sigma)\frac{S_0(y)}{1 - S_0(y)}\,. \tag{3.4}$$

Several authors have discussed the properties of the accelerated failure time model and the proportional hazards model. Lindsey (1998) points out that the eventual disadvantage of the former model —the need to specify the distribution— is compensated by the availability of the hazard function. This function reflects the instantaneous risk of Y whereas $S(y)$ is a cumulative measure. Besides, according to the author, this model is robust under different distributions for heavy censoring. Nardi and Schemper (2003) mention advantages of the parametric model over the Cox model under certain circumstances such as time trends in covariates or parameter values far from zero. In another instance, Hougaard (1999) shows that the Cox model is more adequate to incorporate time-varying covariates, whereas the log linear model is less sensitive when significant covariates are neglected by the model. Similar conclusions are discussed in Hutton and Monaghan (2002) and Frankel and Longmate (2002). Finally, the accelerated failure time model might also be treated as a semi-parametric model if the distribution W in model (3.3) remains unspecified. Studies of Ying, Wei, and Lin (1992) and Park and Wei (2003) deal with this model.

In the remainder of that chapter, we will use the log linear expression in (3.3) when referring to the accelerated failure time model.

3.2 Model, censoring patterns, and observable data

The model we deal with in the sequel is a simple log linear model, that is, for simplicity reasons, we only consider one covariate Z:

$$\ln(Y) = \mu + \beta Z + \sigma W. \tag{3.5}$$

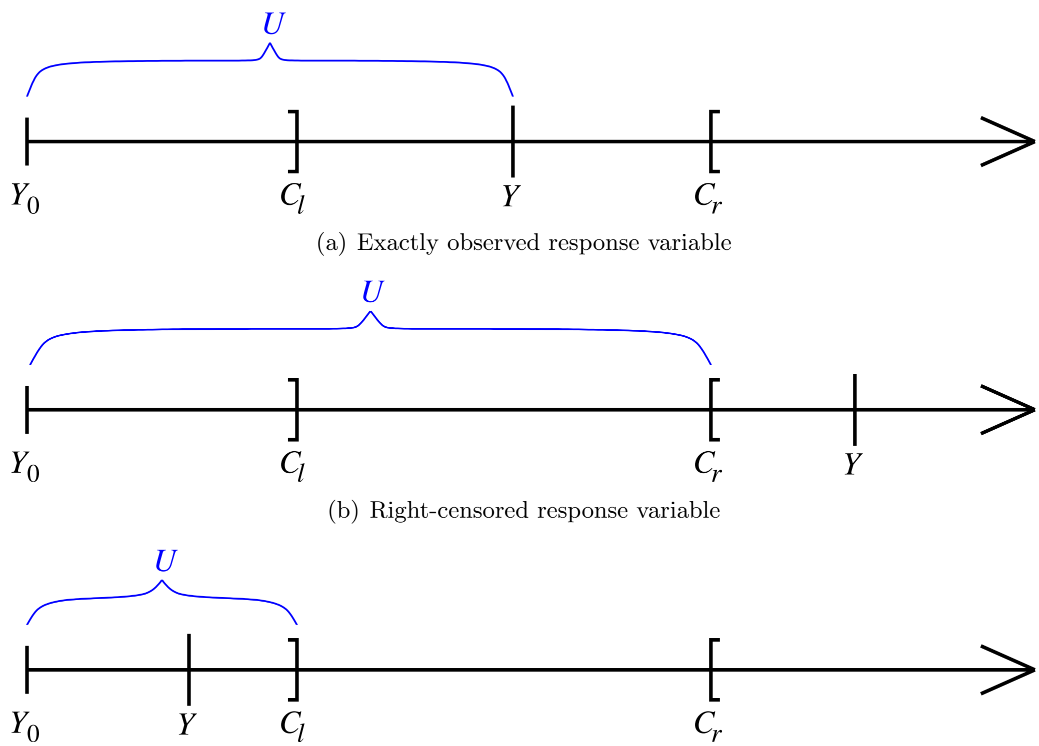
The particularity, which we consider, refers to the censoring of both the response variable Y and the covariate Z. In case of Z, we observe intervals $I_Z = [Z_l, Z_r]$ with $\mathrm{P}(Z \in I_Z) = 1$, which include left-censored data ($Z_l = 0$) and right-censoring ($Z_r = \infty$) as particular cases. The intervals I_Z are assumed closed, but the methodology ahead is easily adapted to the case of semi-open intervals.

Concerning the response variable Y, we distinguish two cases according to the censoring pattern. On one hand, Y may be left- or right-censored, on the other hand, Y is doubly-censored. We shall denote these censoring patterns by case 1 and case 2.

Case 1: Left- and right-censored response variable In this case, Y may be either exactly observed or left- or right-censored. For the sake of simplicity, we do not consider the case of interval-censoring, but generalizations to include this type of censoring are straightforward. However, interval-censoring arises often when the probability for exact observations is equal to zero, whereas left- and right-censored data are observed when also exact observations are possible. An example for interval-censoring is the HIV infection and for left- and right-censoring the moment of AIDS onset as illustrated in Section 2.2.

The three observational patterns are illustrated in the following Figure 3.1, where Y_0 stands for the time origin of Y and U, C_r, and C_l denote the random variables of the observation times, the right-censoring, and the left-censoring times, respectively. The variable U is defined as

$$U = \max\big(C_l, \min(Y, C_r)\big) = \min\big(\max(C_l, Y), C_r\big). \tag{3.6}$$



(a) Exactly observed response variable

(b) Right-censored response variable

(c) Left-censored response variable

Figure 3.1: Distinction of case 1 censoring patterns

To distinguish the different types of censoring, indicator functions have to be defined:

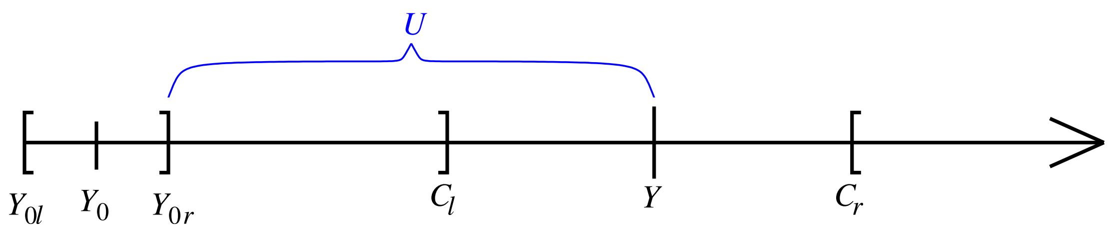
$$\delta_1 = \begin{cases} 1 & Y \text{ is observed exactly} \\ 0 & \text{otherwise} \end{cases},$$
$$\delta_2 = \begin{cases} 1 & Y \text{ is right-censored} \\ 0 & \text{otherwise} \end{cases}. \tag{3.7}$$

That is, $\delta_1 = 1$ and $\delta_2 = 0$ indicate an exact observation, $\delta_1 = 0$ and $\delta_2 = 1$ a right-censored one, and $\delta_1 = \delta_2 = 0$ defines left-censoring. Note that $\delta_1 = \delta_2 = 1$ is not possible.

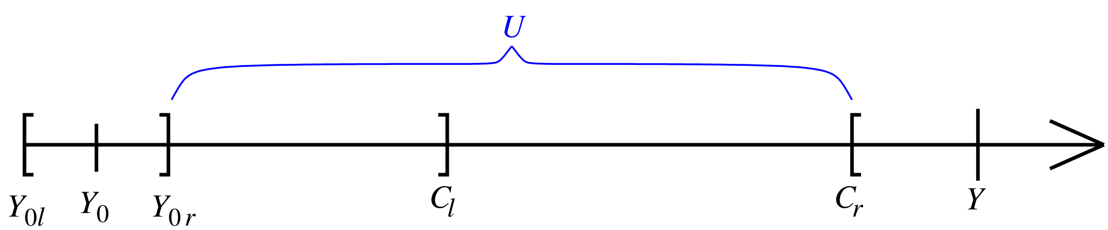
Hence, for case 1, the observation data vector for each subject can be summarized by:

$$(U, Z_l, Z_r, \delta_1, \delta_2). \tag{3.8}$$

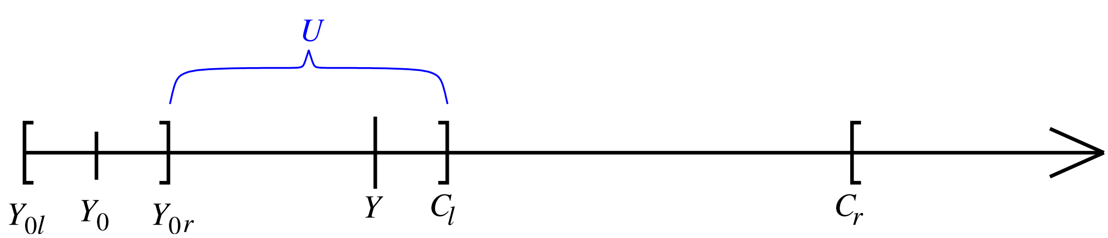
Case 2: Doubly-censored response variable The survival time Y is doubly-censored as illustrated in Figures 3.2(a)–3.2(c), where $[Y_{0l}, Y_{0r}]$ denotes the interval containing the unobserved time origin Y_0. We define the indicator functions δ_1 and δ_2 in the same way as in (3.7), but the variable U cannot be measured from Y_0. Instead, it is measured from time Y_{0r}.



(a) Exactly observed endpoint



(b) Right-censored endpoint



(c) Left-censored endpoint

Figure 3.2: Distinction of case 2 censoring patterns

For case 2, the observation data vector for each individual, including Z, is the following:

$$(U, Y_{0l}, Y_{0r}, Z_l, Z_r, \delta_1, \delta_2). \tag{3.9}$$

Denoting by Δ_{Y_0} the width of the interval $[Y_{0l}, Y_{0r}]$, that is, $\Delta_{Y_0} = Y_{0r} - Y_{0l}$, we have

$$Y \in \begin{cases} [U, U + \Delta_{Y_0}] & \text{if } \delta_1 = 1, \delta_2 = 0 \\ [U, \infty) & \text{if } \delta_1 = 0, \delta_2 = 1 \\ (0, U + \Delta_{Y_0}] & \text{if } \delta_1 = 0, \delta_2 = 0 \end{cases} . \tag{3.10}$$

Given case 2, a right-censored time origin of Y, that is $Y_{0r} = \infty$, implies missing endpoint of Y and therefore a missing value of the response variable. This case shall be discussed in Section 3.6.2, beforehand such data will not be considered for the construction of the likelihood function.

The data from the Hospital Can Ruti, described in Section 2.2, present a particular case of case 2, because the doubly-censored AIDS incubation period is the subsequent time of the time until HIV infection. Hence, the censoring interval $[Z_l, Z_r]$ of the covariate and the interval $[Y_{0l}, Y_{0r}]$ coincide and the observed data vector in (3.9) reduces to $(U, Z_l, Z_r, \delta_1, \delta_2)$ as for case 1.

3.3 Likelihood functions

In all cases, our objective is to determine the maximum likelihood estimate of the unknown parameter vector $\boldsymbol{\theta} = (\mu, \beta, \sigma)'$. For this purpose, we first derive the noninformativity conditions and construct then the likelihood functions. This has to be done explicitly, because, so far, statistical software does not cover regression models with an interval-censored covariate.

3.3.1 Noninformativity conditions

For the construction of the likelihood functions for case 1 and case 2, it shall be assumed that censoring of both the response variable and the covariate is noninformative following the definitions in Section 1.2. Under these conditions, the censoring generating process needs not to be considered in the likelihood functions.

Concerning interval-censoring of the covariate Z, we can summarize the conditions of noninformative censoring for any z, z_l, z_r, and y as follows:

$$f(z|Z_l = z_l, Z_r = z_r) = \frac{f(z)}{\mathrm{P}(z_l \le Z \le z_r)} \mathbb{1}_{\{z \in [z_l, z_r]\}}, \tag{3.11a}$$

$$f(y|Z = z, Z_l = z_l, Z_r = z_r) = f(y|Z = z). \tag{3.11b}$$

Condition (3.11a) implies that the interval $I_Z = [Z_l, Z_r]$ carries no further information about the real unobserved value of Z; rather, Z lies in I_Z. On the other hand, according to condition (3.11b), the response variable does only depend on Z but not on I_Z.

In case of the response variable's censoring, it is assumed that both censoring of the time origin Y_0 and the endpoint does not inform on the unobserved real values. Formally, this can be summarized by the following two expressions for any values y_{0l}, y_{0r}, y, and u. Condition (3.11c) refers to case 1, the following to case 2:

$$f(y|U=u, Z=z) = \begin{cases} \frac{f(y|z)}{\mathrm{P}(Y>u|z)} \mathbb{1}_{\{y>u\}} & \text{if } \delta_1 = 0, \delta_2 = 1 \\ \frac{f(y|z)}{\mathrm{P}(Y<u|z)} \mathbb{1}_{\{y<u\}} & \text{if } \delta_1 = 0, \delta_2 = 0 \end{cases}, \tag{3.11c}$$

$$f(y|Y_{0l} = y_{0l}, Y_{0r} = y_{0r}, U = u, Z = z) =$$

$$\begin{cases} \frac{f(y|z)}{\mathrm{P}(u \le Y \le u+\delta_{Y_0}|z)} \mathbb{1}_{\{y \in [u, u+\delta_{Y_0}]\}} & \text{if } \delta_1 = 1, \delta_2 = 0 \\ \frac{f(y|z)}{\mathrm{P}(Y>u|z)} \mathbb{1}_{\{y>u\}} & \text{if } \delta_1 = 0, \delta_2 = 1 \\ \frac{f(y|z)}{\mathrm{P}(Y<u+\delta_{Y_0}|z)} \mathbb{1}_{\{y<u+\delta_{Y_0}\}} & \text{if } \delta_1 = 0, \delta_2 = 0 \end{cases}, \tag{3.11d}$$

where $\delta_{Y_0} = y_{0r} - y_{0l}$. Note that the case $\delta_1 = 1, \delta_2 = 0$ does not need to be considered in (3.11c) since it describes the exactly observed endpoint of Y. Note that the conditions(3.11c) and (3.11d) use the conditional distribution of Y given Z, as we will need this conditional distribution for the construction of the following likelihood functions.

3.3.2 Construction of the likelihood functions

Herein, the likelihood functions corresponding to model (3.5) are derived for both cases assuming the noninformativity conditions (3.11) hold. The resulting functions do not only depend on the parameter vector $\boldsymbol{\theta} = (\mu, \beta, \sigma)'$, but also on the (unknown) distribution functions F_Z of the interval-censored covariate and —given case 2— F_{Y_0} of the response variable's time origin Y_0. The derivation of the likelihood is done heuristically, in part, resembling the procedure of Commenges (2003) in the case of multi-state models with interval-censored observations.

Case 1: Left- and right-censored response variable

Given the observation vector (3.8), the likelihood contribution of a single individual is defined by the values of the indicator variables δ_1 and δ_2, that describe the censoring of Y. It is the value of either the conditional density function of Y given Z, the conditional survival function, or one minus the survival function. Since Z is not observed exactly, the whole interval $[Z_l, Z_r]$ must be taken into account, which is accomplished by integrating over all values of the interval. The likelihood contribution can then be summarized by the following expression:

$$L_{cont}(\boldsymbol{\theta}, F_Z) = \int_{z_l}^{z_r} f(u|z)^{\delta_1} S(u|z)^{\delta_2} \big(1 - S(u|z)\big)^{(1-\delta_1)(1-\delta_2)} \, dF_Z(z). \tag{3.12}$$

Given the independent observations $(u_i, z_{l_i}, z_{r_i}, \delta_{1_i}, \delta_{2_i})$, $i = 1, \ldots, n$, the resulting likelihood function $L(\boldsymbol{\theta}, F_Z)$ is the product over all likelihood contributions $C_i(\boldsymbol{\theta}, F_Z)$:

$$L(\boldsymbol{\theta}, F_Z) = \prod_{i=1}^{n} \int_{z_{l_i}}^{z_{r_i}} f(u_i|z)^{\delta_{1_i}} S(u_i|z)^{\delta_{2_i}} \big(1 - S(u_i|z)\big)^{(1-\delta_{1_i})(1-\delta_{2_i})} \, dF_Z(z). \tag{3.13}$$

So far, survival models with interval-censored covariates are not considered by statistical software. For this reason, the maximization of $L(\boldsymbol{\theta}, F_Z)$ would be very cumbersome in case of a continuous covariate, at least as long as its distribution is not specified parametrically.

In contrast with that, if the covariate is assumed discrete, the integrals in (3.13) disappear and maximization becomes feasible. For this purpose, let Z be a discrete random variable with support $S = \{s_1, \ldots, s_m\}$, where $s_1 < s_2 < \cdots < s_m$, and the corresponding probabilities $\omega_j = \mathrm{P}(Z = s_j)$, $j = 1, \ldots, m$. Besides, the following indicator variables are defined:

$$\alpha_{ij} = \mathbb{1}_{\{s_j \in [z_{l_i}, z_{r_i}]\}}, \; i = 1, \ldots, n, \; j = 1, \ldots, m. \tag{3.14}$$

That is, the variable α_{ij} is equal to 1, if s_j is an admissible value for Z_i, and zero otherwise. The likelihood function (3.13) has then the following expression, where $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_m)'$:

$$L(\boldsymbol{\theta}, \boldsymbol{\omega}) = \prod_{i=1}^{n} \sum_{j=1}^{m} \alpha_{ij} f(u_i|s_j)^{\delta_{1_i}} S(u_i|s_j)^{\delta_{2_i}} \big(1 - S(u_i|s_j)\big)^{(1-\delta_{1_i})(1-\delta_{2_i})} \omega_j, \tag{3.15}$$

This function resembles the likelihood function (1.7) on page 9 derived by Gómez, Espinal, and Lagakos (2003), who deal with the case of a linear regression model with an interval-censored covariate. The log likelihood function, $l(\boldsymbol{\theta}, \boldsymbol{\omega}) = \ln L(\boldsymbol{\theta}, \boldsymbol{\omega})$, is equal to

$$l(\boldsymbol{\theta}, \boldsymbol{\omega}) = \sum_{i=1}^{n} \ln \Big(\sum_{j=1}^{m} \alpha_{ij} f(u_i|s_j)^{\delta_{1_i}} S(u_i|s_j)^{\delta_{2_i}} \big(1 - S(u_i|s_j)\big)^{(1-\delta_{1_i})(1-\delta_{2_i})} \omega_j \Big). \tag{3.16}$$

Our main statistical goal is the maximum likelihood estimation of $\boldsymbol{\theta}$ in the presence of the nuisance parameter vector $\boldsymbol{\omega}$. Hence, we are left with the computational burden of maximizing (3.16) subject to the following constraints:

$$\sum_{j=1}^{m} \omega_j = 1, \tag{3.17a}$$

$$\omega_j \geq 0, \; j = 1, \ldots, m. \tag{3.17b}$$

Following the notation in Section 1.3.1 on page 12, a nonlinear constrained optimization problem is to be solved with objective function $l(\boldsymbol{\theta}, \boldsymbol{\omega})$, equality constraint (3.17a), and inequality constraints (3.17b).

Case 2: Doubly-censored response variable

Looking at the intervals in (3.10), one might tend to use these as observed intervals for Y and proceed as for case 1 including the case of interval censoring in the likelihood contributions (3.12). However, as mentioned before in Section 1.1.3 on page 8, this procedure would not be valid since the independence between the observation times and event time distribution would be lost by this transformation. An exception is the case of a uniform distribution for Y_0 (De Gruttola and Lagakos 1989).

Therefore, to address the general case, the likelihood function must account for the distribution of the chronological times of the time origin Y_0. Denoting by F_{Y_0} the distribution function of Y_0 and supposing Z was observed exactly, the likelihood contribution of an individual would be equal to the following expression:

$$L_{cont}(\boldsymbol{\theta}, F_{Y_0}) = \int_{y_{0l}}^{y_{0r}} f\big(u(d)\big)^{\delta_1} S\big(u(d)\big)^{\delta_2} \big(1 - S\big(u(d)\big)\big)^{(1-\delta_1)(1-\delta_2)} \, dF_{Y_0}(d),$$

where $u(d)$ is equal to $u + y_{0r} - d$ as illustrated in Figure 3.3 on the following page. That is, with every possible value $d \in [Y_{0l}, Y_{0r}]$, the value of U changes.

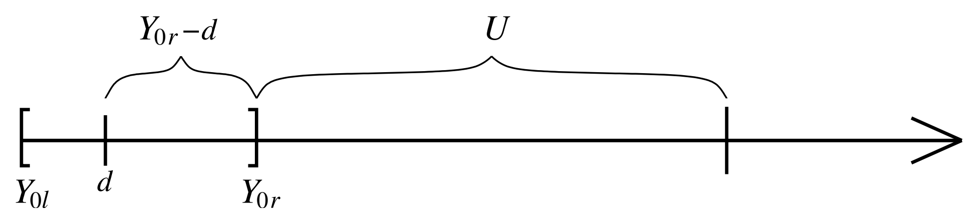


Figure 3.3: Calculation scheme for a doubly-censored response variable

However, being Z actually interval-censored, the likelihood contribution for case 2 is given by

$$L_{cont}(\boldsymbol{\theta}, F_{Y_0}, F_Z) = \int_{y_{0l}}^{y_{0r}} \int_{z_l}^{z_r} f\big(u(d)|z\big)^{\delta_1} S\big(u(d)|z\big)^{\delta_2} \big(1 - S\big(u(d)|z\big)\big)^{(1-\delta_1)(1-\delta_2)} \, dF_Z(z) dF_{Y_0}(d).$$

The resulting likelihood function is the product over the likelihood contributions $C_i(\boldsymbol{\theta}, F_{Y_0}, F_Z)$ given the independent observations $(u_i, y_{0l_i}, y_{0r_i}, z_{l_i}, z_{r_i}, \delta_{1_i}, \delta_{2_i})$, $i = 1, \ldots, n$:

$$L(\boldsymbol{\theta}, F_{Y_0}, F_Z) = \prod_{i=1}^{n} \int_{y_{0l_i}}^{y_{0r_i}} \int_{z_{l_i}}^{z_{r_i}} f\big(u_i(d)|z\big)^{\delta_{1_i}} S\big(u_i(d)|z\big)^{\delta_{2_i}} \big(1 - S\big(u_i(d)|z\big)\big)^{(1-\delta_{1_i})(1-\delta_{2_i})} \, dF_Z(z) dF_{Y_0}(d). \quad (3.18)$$

A mathematically more tractable expression than (3.18) is obtained if both the distribution of Z and of Y_0 are assumed discrete. Let Z be defined as in case 1 and let Y_0 be a discrete random variable with support $D = \{d_1, \ldots, d_k\}$, where $d_1 < d_2 < \cdots < d_k$, and the corresponding probabilities $\nu_l = \mathrm{P}(Y_0 = d_l)$, $l = 1, \ldots, k$. Analogously to the variables α_{ij} in (3.14), the following indicator variables are defined:

$$\gamma_{il} = \mathbb{1}_{\{d_l \in [y_{0l_i}, y_{0r_i}]\}}, \; i = 1, \ldots, n, \; l = 1, \ldots, k.$$

That is, γ_{il} is equal to 1, if d_l is an admissible value for Y_{0_i}, and zero otherwise. For this discrete case, the likelihood function has the following expression:

$$L(\boldsymbol{\theta}, \boldsymbol{\nu}, \boldsymbol{\omega}) = \prod_{i=1}^{n} \sum_{l=1}^{k} \sum_{j=1}^{m} \gamma_{il} \alpha_{ij} f\big(u_i(d_l)|s_j\big)^{\delta_{1_i}} S\big(u_i(d_l)|s_j\big)^{\delta_{2_i}} \big(1 - S\big(u_i(d_l)|s_j\big)\big)^{(1-\delta_{1_i})(1-\delta_{2_i})} \omega_j \nu_l, \quad (3.19)$$

where $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_k)'$ and $u_i(d_l) = u_i + y_{0r_i} - d_l$. The maximization of this function or its respective log likelihood function is subject to both the restrictions (3.17) on $\boldsymbol{\omega}$ and the restrictions (3.20) on $\boldsymbol{\nu}$:

$$\sum_{l=1}^{k} \nu_l = 1, \qquad \nu_l \geq 0, \; l = 1, \ldots, k. \quad (3.20)$$

Likelihood function for data from Can Ruti

As mentioned before, for this particular case, the observed data can be summarized by vector (3.8) of case 1. However, the construction of the likelihood function, has to account for the fact that the interval-censored endpoint of the covariate Z, that is, HIV infection, coincides with the time origin of the response variable Y. For this reason and given the independent observations $(u_i, z_{l_i}, z_{r_i}, \delta_{1_i}, \delta_{2_i})$, $i = 1, \ldots, n$, the likelihood function (3.18) reduces to the following expression (3.21). Only one integral is needed since any possible value of $Z \in [Z_l, Z_r]$ determines the value of the variable U:

$$L(\boldsymbol{\theta}, F_Z) = \prod_{i=1}^{n} \int_{z_{l_i}}^{z_{r_i}} f\big(u_i(z)|z\big)^{\delta_{1_i}} S\big(u_i(z)|z\big)^{\delta_{2_i}} \big(1 - S\big(u_i(z)|z\big)\big)^{(1-\delta_{1_i})(1-\delta_{2_i})} \, dF_Z(z), \quad (3.21)$$

where $u_i(z) = u_i + z_{r_i} - z$. The corresponding discrete version is similar to (3.15) with indicator variables α_{ij} are defined as in (3.14). The difference lies in that the values u_i change with each considered value of Z, that is, $u_i(s_j) = u_i + z_{r_i} - s_j$:

$$L(\boldsymbol{\theta}, \boldsymbol{\omega}) = \prod_{i=1}^{n} \sum_{j=1}^{m} \alpha_{ij} f\big(u_i(s_j)|s_j\big)^{\delta_{1_i}} S\big(u_i(s_j)|s_j\big)^{\delta_{2_i}} \big(1 - S\big(u_i(s_j)|s_j\big)\big)^{(1-\delta_{1_i})(1-\delta_{2_i})} \omega_j. \quad (3.22)$$

Hence, to obtain the maximum likelihood estimates $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\omega}}$, we have to solve the nonlinear constrained optimization problem with objective function (3.22) and constraints (3.17).

In the sequel, we consider two different parametric choices for the error distribution W. Given the difficulties to maximize the (log) likelihood functions when Z is continuous, parametrically specified or not, only the case of a discrete covariate is considered. That is, we focus the work on maximizing the respective log likelihood functions subject to the present constraints.

3.4 Parametric choices

Herein, the specific forms of the above likelihood functions are addressed for two common parametric choices of model (3.5): the Weibull and the log logistic regression model. As mentioned in Section 3.1, the Weibull model is the only parametric choice which is equivalent to the proportional hazards model (1.5), whereas the log logistic model is equivalent to the proportional odds model (3.4).

For both models, the different log likelihood functions according to the different censoring patterns are presented. In the next section, we will treat the maximization of these functions.

3.4.1 The Weibull regression model

The Weibull distribution

The density and survival functions of a Weibull distributed random variable Y with scale or location parameter $\lambda\,(>0)$ and shape parameter $\alpha\,(>0)$, $Y \sim \mathcal{W}(\lambda, \alpha)$, are given by

$$
\begin{aligned}
f(y) &= \alpha\lambda y^{\alpha-1}\exp(-\lambda y^{\alpha}),\\
S(y) &= \exp(-\lambda y^{\alpha}).
\end{aligned}
$$

The Weibull distribution is a flexible distribution since it allows for a decreasing hazard function ($\alpha < 1$), an increasing hazard function ($\alpha > 1$) and constant hazards. The latter case is true for $\alpha = 1$; then, Y follows an exponential distribution with parameter λ.

Relation with the log linear model

A Weibull distribution $Y \sim \mathcal{W}(\lambda, \alpha)$ can also be described by means of a log linear model with parameters μ and σ, where the error distribution W is the standard Gumbel distribution:

$$\ln(Y) = \mu + \sigma W.$$

The relation between the parameters is: $\alpha = 1/\sigma$ and $\lambda = \exp(-\mu/\sigma)$. The density and survival functions of the standard Gumbel or extreme value distribution are given by

$$\begin{aligned} f(w) &= \exp(w - e^w), \\ S(w) &= \exp(-e^w). \end{aligned} \tag{3.23}$$

The expression as a log linear model has the advantage that covariates can be incorporated such as in model (3.5). The parameters of the Weibull distribution depend then on the value of the covariate Z. Whereas the shape parameter $\alpha\,(= 1/\sigma)$ is the same for all conditional survival times Y given Z, the location parameter changes with Z: $\lambda(z) = \exp\big(-(\mu+\beta z)/\sigma\big)$. Due to the invariance property, given the maximum likelihood estimates $(\hat{\mu}, \hat{\beta}, \hat{\sigma})$, the maximum likelihood estimates $\hat{\alpha}$ and $\hat{\lambda}(z)$ are easily obtained applying the corresponding transformations.

Using the Weibull regression model, the term $\exp(-\beta/\sigma)$ corresponds to the relative risk and $\exp(\beta)$ to the acceleration factor when comparing two individuals, whose covariate values differ by one unit. The interpretation of these terms implies that augmenting the covariate by one, the risk of dying increases/decreases ($\beta < 0/\beta > 0$) by the factor $\exp(-\beta/\sigma)$, whereas the median time until the event of interest is decreased/increased ($\beta < 0/\beta > 0$) by the factor $\exp(\beta)$.

Log likelihood functions

In case of the Weibull regression model, the density and survival functions of Y given Z in the above likelihood functions (3.15), (3.19), and (3.22) have to be substituted by the respective functions (3.23) of the standard Gumbel distribution. Using the index $_Y$ for the Weibull distributed survival time Y and $_W$ for the Gumbel distributed error of the model (3.5), we have the following relations (Klein and Moeschberger 1997, Chap. 12)

$$\begin{aligned} f_Y(y) &= \tfrac{1}{\sigma} f_W\big(\tfrac{\ln(y)-\mu-\beta z}{\sigma}\big) = \tfrac{1}{\sigma}\exp\big(\tfrac{\ln(y)-\mu-\beta z}{\sigma} - e^{\frac{1}{\sigma}(\ln(y)-\mu-\beta z)}\big), \\ S_Y(y) &= S_W\big(\tfrac{\ln(y)-\mu-\beta z}{\sigma}\big) = \exp\big(-e^{\frac{1}{\sigma}(\ln(y)-\mu-\beta z)}\big). \end{aligned} \tag{3.24}$$

Using the density and survival functions of the extreme value distribution in (3.24), the log likelihood functions for case 1, case 2, and the data from Can Ruti can be written as follows:

Case 1: Left- and right-censored response variable

$$\begin{aligned} l(\boldsymbol{\theta},\boldsymbol{\omega}) = \sum_{i=1}^{n} \Bigg[& \delta_{1_i} \ln \Big(\sum_{j=1}^{m} \frac{\alpha_{ij}}{\sigma} \exp\big(\tfrac{\ln(u_i)-\mu-\beta s_j}{\sigma} - e^{\frac{1}{\sigma}(\ln(u_i)-\mu-\beta s_j)}\big)\omega_j \Big) \\ & + \delta_{2_i} \ln \Big(\sum_{j=1}^{m} \alpha_{ij} \exp\big(- e^{\frac{1}{\sigma}(\ln(u_i)-\mu-\beta s_j)}\big)\omega_j \Big) \\ & + (1-\delta_{1_i})(1-\delta_{2_i}) \ln \Big(\sum_{j=1}^{m} \alpha_{ij} \Big(1-\exp\big(- e^{\frac{1}{\sigma}(\ln(u_i)-\mu-\beta s_j)}\big)\Big)\omega_j \Big) \Bigg]. \end{aligned} \tag{3.25}$$

Note that the transformation from (3.16) to (3.25) is only valid, because both δ_1 and δ_2, as well as their sum are equal to either zero or one. This holds also for the following expressions.

Case 2: Doubly-censored response variable

$$\begin{aligned} l(\boldsymbol{\theta},\boldsymbol{\nu},\boldsymbol{\omega}) = \sum_{i=1}^{n} \Bigg[& \delta_{1_i} \ln \Big(\sum_{l=1}^{k}\sum_{j=1}^{m} \frac{\gamma_{il}\alpha_{ij}}{\sigma} \exp\big(\tfrac{\ln(u_i(d_l))-\mu-\beta s_j}{\sigma} - e^{\frac{1}{\sigma}(\ln(u_i(d_l))-\mu-\beta s_j)}\big)\omega_j\nu_l \Big) \\ & + \delta_{2_i} \ln \Big(\sum_{l=1}^{k}\sum_{j=1}^{m} \gamma_{il}\alpha_{ij} \exp\big(- e^{\frac{1}{\sigma}(\ln(u_i(d_l))-\mu-\beta s_j)}\big)\omega_j\nu_l \Big) \\ & + (1-\delta_{1_i})(1-\delta_{2_i}) \ln \Big(\sum_{l=1}^{k}\sum_{j=1}^{m} \gamma_{il}\alpha_{ij} \Big(1-\exp\big(- e^{\frac{1}{\sigma}(\ln(u_i(d_l))-\mu-\beta s_j)}\big)\Big)\omega_j\nu_l \Big) \Bigg]. \end{aligned} \tag{3.26}$$

Log likelihood function for data from Can Ruti

$$\begin{aligned} l(\boldsymbol{\theta},\boldsymbol{\omega}) = \sum_{i=1}^{n} \Bigg[& \delta_{1_i} \ln \Big(\sum_{j=1}^{m} \frac{\alpha_{ij}}{\sigma} \exp\big(\tfrac{\ln(u_i(s_j))-\mu-\beta s_j}{\sigma} - e^{\frac{1}{\sigma}(\ln(u_i(s_j))-\mu-\beta s_j)}\big)\omega_j \Big) \\ & + \delta_{2_i} \ln \Big(\sum_{j=1}^{m} \alpha_{ij} \exp\big(- e^{\frac{1}{\sigma}(\ln(u_i(s_j))-\mu-\beta s_j)}\big)\omega_j \Big) \\ & + (1-\delta_{1_i})(1-\delta_{2_i}) \ln \Big(\sum_{j=1}^{m} \alpha_{ij} \Big(1-\exp\big(- e^{\frac{1}{\sigma}(\ln(u_i(s_j))-\mu-\beta s_j)}\big)\Big)\omega_j \Big) \Bigg]. \end{aligned} \tag{3.27}$$

Thus, to obtain the maximum likelihood estimates for $\boldsymbol{\theta}$, $\boldsymbol{\omega}$, and $\boldsymbol{\nu}$, the previous log likelihood functions (3.25)–(3.27) must be maximized subject to the restrictions (3.17) on $\boldsymbol{\omega}$ and, for case 2, the restrictions (3.20) on $\boldsymbol{\nu}$. Besides, the model parameter σ must be forced to be positive since it is equivalent to the inverse of the Weibull distribution parameter α.

3.4.2 The log logistic regression model

The log logistic distribution

A random variable Y follows a log logistic distribution if the distribution of its natural logarithm is a logistic distribution. The shape of this distribution is similar to the one of the log normal distribution, but preferable in terms of mathematical tractability since its distribution function has a closed expression. As the one of the log normal distribution, its hazard function is hump-shaped, that is, it increases initially and then decreases. Its density and survival functions with parameters α and λ, both strictly positive, are given by

$$f(y) = \frac{\lambda \alpha y^{\alpha-1}}{(1+\lambda y^{\alpha})^2},$$
$$S(y) = \frac{1}{1+\lambda y^{\alpha}}.$$

Relation with the log linear model

A log logistic distribution $Y \sim \log \mathcal{L}(\lambda, \alpha)$ can also be described by means of a log linear model with parameters μ and σ, where the error distribution W is the standard logistic distribution:

$$\ln(Y) = \mu + \sigma W.$$

As for the Weibull model, the relation between the parameters is: $\alpha = 1/\sigma$ and $\lambda = \exp(-\mu/\sigma)$. The density and survival function of the standard logistic distribution are given by

$$\begin{aligned} f(w) &= \frac{e^w}{(1+e^w)^2}, \\ S(w) &= \frac{1}{1+e^w}. \end{aligned} \tag{3.28}$$

Covariates can be incorporated in the log linear model such as in model (3.5). Then, whereas the parameter $\alpha\,(=1/\sigma)$ is the same for all conditional survival times Y given Z, the parameter λ depends on Z: $\lambda(z) = \exp(-(\mu+\beta z)/\sigma)$. Due to the invariance property, given the maximum likelihood estimates $(\hat{\mu}, \hat{\beta}, \hat{\sigma})$, the maximum likelihood estimates $\hat{\alpha}$ and $\hat{\lambda}(z)$ are obtained using the corresponding transformations.

Using the log logistic regression model, the term $\exp(-\beta/\sigma)$ corresponds to the relative odds and $\exp(\beta)$ to the acceleration factor when comparing two individuals, whose covariate values differ by one. The interpretation of these terms implies that augmenting the covariate one unit, the odds of dying increases/decreases ($\beta < 0/\beta > 0$) by the factor $\exp(-\beta/\sigma)$, whereas the median time until the event of interest is decreased/increased ($\beta < 0/\beta > 0$) by the factor $\exp(\beta)$.

Log likelihood functions

In case of the log logistic regression model, the density and survival functions of Y given Z in the above likelihood functions (3.15), (3.19), and (3.22) have to be replaced by the respective functions (3.28) of the standard logistic distribution. Using the index $_Y$ for the log logistic distributed survival time Y and $_W$ for the logistic distributed error of the model (3.5), we have (Klein and Moeschberger 1997, Chap. 12):

$$\begin{aligned} f_Y(y) &= \tfrac{1}{\sigma} f_W\left(\tfrac{\ln(y)-\mu-\beta z}{\sigma}\right) = \frac{\exp(\frac{1}{\sigma}(\ln(y)-\mu-\beta z))}{\sigma\big(1+\exp(\frac{1}{\sigma}(\ln(y)-\mu-\beta z))\big)^2}, \\ S_Y(y) &= S_W\left(\tfrac{\ln(y)-\mu-\beta z}{\sigma}\right) = \frac{1}{1+\exp\big(\frac{1}{\sigma}(\ln(y)-\mu-\beta z)\big)}. \end{aligned} \tag{3.29}$$

Using the density and survival function of the standard logistic distribution in (3.29), the log likelihood functions for case 1, case 2, and the data from Can Ruti can be written as follows:

Case 1: Left- and right-censored response variable

$$\begin{aligned} l(\boldsymbol{\theta},\boldsymbol{\omega}) = \sum_{i=1}^{n} \Bigg[& \delta_{1_i} \ln\Big(\sum_{j=1}^{m} \frac{\alpha_{ij}}{\sigma} \frac{\exp(\frac{1}{\sigma}(\ln(u_i)-\mu-\beta s_j))}{\big(1+\exp(\frac{1}{\sigma}(\ln(u_i)-\mu-\beta s_j))\big)^2} \omega_j \Big) \\ & + \delta_{2_i} \ln\Big(\sum_{j=1}^{m} \frac{\alpha_{ij}}{1+\exp\big(\frac{1}{\sigma}(\ln(u_i)-\mu-\beta s_j)\big)} \omega_j \Big) \\ +(1-\delta_{1_i})(1-\delta_{2_i}) & \ln\Big(\sum_{j=1}^{m} \alpha_{ij}\Big(1-\frac{1}{1+\exp\big(\frac{1}{\sigma}(\ln(u_i)-\mu-\beta s_j)\big)}\Big)\omega_j \Big) \Bigg]. \end{aligned} \tag{3.30}$$

Case 2: Doubly-censored response variable

$$\begin{aligned} l(\boldsymbol{\theta},\boldsymbol{\nu},\boldsymbol{\omega}) = \sum_{i=1}^{n} \Bigg[& \delta_{1_i} \ln\Big(\sum_{l=1}^{k}\sum_{j=1}^{m} \frac{\gamma_{il}\alpha_{ij}}{\sigma} \frac{\exp(\frac{1}{\sigma}(\ln(u_i(d_l))-\mu-\beta s_j))}{\big(1+\exp(\frac{1}{\sigma}(\ln(u_i(d_l))-\mu-\beta s_j))\big)^2} \omega_j \nu_l \Big) \\ & + \delta_{2_i} \ln\Big(\sum_{l=1}^{k}\sum_{j=1}^{m} \frac{\gamma_{il}\alpha_{ij}}{1+\exp\big(\frac{1}{\sigma}(\ln(u_i(d_l))-\mu-\beta s_j)\big)} \omega_j \nu_l \Big) \\ +(1-\delta_{1_i})(1-\delta_{2_i}) & \ln\Big(\sum_{l=1}^{k}\sum_{j=1}^{m} \gamma_{il}\alpha_{ij}\Big(1-\frac{1}{1+\exp\big(\frac{1}{\sigma}(\ln(u_i(d_l))-\mu-\beta s_j)\big)}\Big)\omega_j \nu_l \Big) \Bigg]. \end{aligned} \tag{3.31}$$

Log likelihood function for data from Can Ruti

$$
\begin{aligned}
l(\boldsymbol{\theta},\boldsymbol{\omega}) = \sum_{i=1}^{n} \Bigg[\delta_{1_i} \ln\Big(\sum_{j=1}^{m} \frac{\alpha_{ij}}{\sigma} \frac{\exp(\frac{1}{\sigma}(\ln(u_i(s_j)) - \mu - \beta s_j))}{\big(1+\exp(\frac{1}{\sigma}(\ln(u_i(s_j)) - \mu - \beta s_j))\big)^2} \omega_j \Big) \\
+\delta_{2_i} \ln\Big(\sum_{j=1}^{m} \frac{\alpha_{ij}}{1+\exp\left(\frac{1}{\sigma}(\ln(u_i(s_j)) - \mu - \beta s_j)\right)} \omega_j \Big) \\
+(1-\delta_{1_i})(1-\delta_{2_i}) \ln\Big(\sum_{j=1}^{m} \alpha_{ij} \Big(1 - \frac{1}{1+\exp\left(\frac{1}{\sigma}(\ln(u_i(s_j)) - \mu - \beta s_j)\right)}\Big) \omega_j \Big) \Bigg].
\end{aligned} \tag{3.32}
$$

As in case of the Weibull regression model, the maximization of each of the log likelihood functions (3.30)–(3.32) is subject to the restrictions (3.17) on $\boldsymbol{\omega}$, and the maximization of (3.31) is also restricted to the constraints (3.20) on $\boldsymbol{\nu}$. As well, the parameter σ must be positive given its relation with the parameter α of the underlying log logistic distributions.

3.5 Simultaneous maximum likelihood estimation

Once derived the expression of the corresponding (log) likelihood function, the objective consists in estimating both the model parameters $\boldsymbol{\theta} = (\mu, \beta, \sigma)'$ and the unknown distribution function of the discrete interval-censored covariate Z characterized by $\boldsymbol{\omega}$. Looking at the log likelihood functions (3.25)–(3.27) of the Weibull regression model and (3.30)–(3.32) of the log logistic model, it is obvious that the maximum likelihood estimators cannot be obtained analytically, but that maximization has to be carried out numerically.

For ease of notation, in the sequel, we denote by $(\hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\omega}}_n)$ the joint maximum likelihood estimators, whether they correspond to case 1 or case 2, which also includes $\hat{\boldsymbol{\nu}}_n$.

3.5.1 General inference procedure

To make inference on the possible effect of the covariate Z on the response Y, we need not only to estimate the model parameters, but also to calculate the corresponding confidence intervals. The maximum likelihood estimates $(\hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\omega}}_n)$ are obtained by the use of AMPL in combination with the NEOS solvers as described in detail in the following Section 3.5.2.

For the computation of the confidence intervals, another software package has to be used, because AMPL does not provide an estimation of the parameters' variances. For instance, the mathematical software Maple is an appropriate tool for this. The computation takes advantage of the asymptotic properties of maximum likelihood estimators: on one hand, the variances can be estimated by inverting the observed Fisher information matrix, on the other hand, the estimators are asymptotically normal distributed (Cox and Hinkley 1974; Lindsey 1996).

Applying the invariance property, the estimation of the relative risk (odds ratio in case of the log logistic regression models) and the acceleration factor, both introduced in Section 3.1, is straightforward:

$$\widehat{\mathrm{RR}}_n = \widehat{\mathrm{OR}}_n = \exp(-\hat{\beta}_n Z/\hat{\sigma}_n), \tag{3.33a}$$

$$\widehat{\mathrm{AF}}_n = \exp(\hat{\beta}_n Z). \tag{3.33b}$$

Note that if Z is a dichotomic variable or when comparing two individuals with covariate values $z+1$ and z, both terms are estimated by $\exp(-\hat{\beta}_n/\hat{\sigma}_n)$ and $\exp(\hat{\beta}_n)$, respectively.

Given the variance of $\hat{\boldsymbol{\theta}}$, we can also calculate the confidence intervals of the relative risk (or odds ratio), and the acceleration factor. Whereas the latter is a mere transformation of the confidence interval of β following (3.33b), for the former, we need to apply the delta method (Schervish 1995, Sec. 7.1) to obtain the estimated variances; see Section B.2 for its derivation in case of a dichotomic covariate Z.

3.5.2 Maximization of the log likelihood

A possible approach to the maximization problems would be to adapt the algorithm of Gómez et al. (2003), described in Section 1.1.4 on page 9, to the given log likelihood functions. However, instead of alternating the two iterative steps of this algorithm and supported by the computing facilities nowadays, we propose the simultaneous maximization of the respective log likelihood functions with respect to $(\boldsymbol{\theta}, \boldsymbol{\omega})$. For the solution of this nonlinear constrained maximization problems, advantage is taken of the methods from the area of operations research.

As an illustration, consider the log likelihood function (3.25) corresponding to the Weibull regression model with a left- and right-censored response variable. Its maximization is subject to the equality constraint $\sum_{j=1}^{m} \omega_j = 1$ and the inequality constraints $\omega_j \geq 0$, $j = 1, \ldots, m$ and $\sigma \geq 0$. Actually, the Weibull distribution requires $\sigma > 0$, but algorithms for optimization problems cannot handle strict inequality constraints because of limited exactness to distinguish reals. However, if that is a problem in practice, the inequality constraint can slightly be modified to $\sigma \geq \epsilon > 0$.

In Section 1.3.2, important tools for the solution of optimization problems have been presented, whose use is recommended to determine $(\hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\omega}}_n)$. The mathematical programming language AMPL is an appropriate tool, since optimization problems written in the AMPL code can be solved by a broad range of specific solvers available at the NEOS server. Hence, apart from programming the corresponding files in AMPL, it is also essential to choose an adequate solver to maximize the log likelihood functions.

For the given maximization problems, we recommend the solver SNOPT. As pointed out on page 14, this solver is suitable for large nonlinearly constrained optimization problems with a modest number of degrees of freedom, especially if the objective function and their gradients are costly to evaluate. The log likelihood functions above have these characteristics: they consist of n summands, each of which being the logarithm of sums itself. Hence, the gradients of $l(\boldsymbol{\theta}, \boldsymbol{\omega})$ are sums of fractions whose denominators consist of sums, too. Our practical experience has confirmed that SNOPT is an adequate choice to handle these maximization problems.

Specification of Z

If the covariate Z is discrete, the maximization procedure requires a specification of the range and values of the covariate's support to be considered. For example, should $S = \{s_1, \ldots, s_m\}$ be a very fine grid or just the opposite? In their work on the nonparametric estimation of a doubly-censored variable's distribution function, De Gruttola and Lagakos (1989) address this aspect, too. If the grid of possible values is chosen too fine, the maximization procedure could be slowed down and the estimation of $\boldsymbol{\omega}$ could furnish nonunique results. On the other hand, if the values are grouped too coarsely, the estimation might not be able to describe well the underlying shape of F_Z. One possibility is to choose the observed times $\{Z_{l_i}\} \cup \{Z_{r_i}\}$, $i = 1, \ldots, n$, to avoid a too fine grid of values $s_j \in S$ and, hence, a too big number of unknown parameters ω_j. This is justified by the fact that any ω_j will be estimated equal to zero, if the corresponding value s_j is neither a left- nor a right-endpoint of any of the observed intervals. Concerning this aspect, the use of a penalized likelihood function could be an interesting aspect of future research, see, for example, Joly and Commenges (1999).

In case right-censored values of the covariate are present, a maximum value s_m has to be determined. If the largest right-censored value $z_l^* = \max\{Z_{l_i} | Z_{r_i} = \infty, i = 1, \ldots, n\}$ does not exceed the largest finite right-endpoint $z_r^* = \max\{Z_{r_i} | Z_{r_i} < \infty, i = 1, \ldots, n\}$, then choose the value $s_m = z_r^*$. A smaller value cannot be chosen and even if $s_m > z_r^*$ is set, the maximization will yield $\hat{F}_Z(z_r^*) = 1$. That is, no positive probability mass will be put on any value beyond z_r^*. On the other hand, if $z_l^* > z_r^*$, we have $\hat{F}_Z(z_r^*) < 1$ and $\hat{\boldsymbol{\omega}}_n$ will not be uniquely defined beyond z_l^*. The remaining probability $1 - \hat{F}(z_r^*)$ will be distributed equally on the interval $[z_l^*, s_m]$, whichever s_m is chosen. Hence, it is important to check whether $z_l^* > z_r^*$ to correctly interpret $\hat{\mathrm{P}}([z_l^*, s_m])$. This is similar to the Kaplan-Meier estimator for right-censored data: it is not defined after the largest right-censored observation if that exceeds the largest uncensored observation.

Multiple maxima of the objective function

Even though the solver SNOPT is an adequate tool to maximize the given log likelihood functions, it cannot be ruled out that the obtained joint maximum likelihood estimator is rather a local than the desired global maximum of $l(\boldsymbol{\theta}, \boldsymbol{\omega})$. This is a characteristic of any solver of nonlinear constrained optimization problems, since no efficient algorithm exists by now which could perform this task.[1]

[1] NEOS guide. URL: `http://www-fp.mcs.anl.gov/otc/Guide` [June 2004]

The problem of multiple roots of the likelihood has been addressed by several authors, see for example, Barnett (1966) and Pewsey (2000). Lindsey (1996, Chap. 3) states that this is not a problem, because there might be several nearly equally likely solutions to the maximization problem. However, this refers to the point estimation of $(\boldsymbol{\theta}, \boldsymbol{\omega})$, whereas the calculation of the corresponding confidence intervals is based on the asymptotic properties of the maximum likelihood estimators. These are valid for the global maximizer of the likelihood function.

One approach to overcome this possible problem shall be mentioned in Section 8.2. However, this is not a particular characteristic of simultaneous maximization. The same problem cannot be ruled neither with other algorithms, such as the EM algorithm.

3.5.3 Estimation of the variance of the model parameters' estimators

As mentioned before, the variance estimation cannot be carried out with AMPL and, consequently, the use of another software package is necessary. One possible tool is the mathematical software package Maple; see, for example, Heck (2003).

In order to estimate the variances of $\boldsymbol{\theta}$ by means of the Fisher information, regularity conditions concerning the estimation procedure must hold. These guarantee the differentiability of the log likelihood function with respect to the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\omega}$ within the corresponding parameters spaces. In particular, one condition is that the true parameters must lie in the interior of the parameter space. However, in case of $\boldsymbol{\omega}$ and $\boldsymbol{\nu}$, these are equal to $[0, 1]^m$ and $[0, 1]^k$, respectively, and hence are closed. As a matter of fact, the simultaneous maximization of the log likelihood functions with respect to ($\boldsymbol{\theta}$,$\boldsymbol{\omega}$), usually yields a considerable number of zero components in $\hat{\boldsymbol{\omega}}_n$. These values lie on the boundary of the parameter space, where the corresponding gradients of the log likelihood function are not equal to zero.

Fay (1996) describes a similar problem in the context of interval-censored data under the grouped continuous model when applying a score test to the model parameters. In order to avoid that the nuisance parameters' estimates approach the boundaries of the parameter space, he reduces the number of nuisance parameters in a first step. This is done in such a way that none of the corresponding estimates approach these boundaries when the score test is applied in the second step.

Here, we are mainly concerned with the inference about $\boldsymbol{\theta}$. For this reason, we base the estimation of the parameters' variances on a reduced version of the likelihood. Assume that the parameter space is reduced to those components with estimates $\hat{\omega}_j > 0$ and we denote by $\boldsymbol{\omega}^*$ the corresponding subvector of $\boldsymbol{\omega}$. Following this notation, we denote by m^* the dimension of $\boldsymbol{\omega}^*$, by s_j^* the values of the support of Z with $\hat{\omega}_j > 0$, and define the indicator variables $\alpha_{ij}^* = \mathbb{1}_{\{s_j^* \in [Z_{l_i}, Z_{r_i}]\}}$, $i = 1, \ldots, n$, $j = 1, \ldots, m^*$. Then, the proposed reduced version of the likelihood function $L^*(\boldsymbol{\theta}, \boldsymbol{\omega}^*)$ has the same expression as the original one, but m, s_j and α_{ij} are replaced by m^*, s_j^* and α_{ij}^*, respectively. For example, consider the likelihood function (3.15) on page 46 in case of a left- and right-censored response variable. The corresponding reduced version of the likelihood function is the following:

$$L^*(\boldsymbol{\theta}, \boldsymbol{\omega}^*) = \prod_{i=1}^{n} \sum_{j=1}^{m^*} \alpha_{ij}^* f(u_i|s_j^*)^{\delta_{1i}} S(u_i|s_j^*)^{\delta_{2i}} \big(1 - S(u_i|s_j^*)\big)^{(1-\delta_{1_i})(1-\delta_{2_i})} \omega_j^*,$$

It is straightforward to prove that, due to the nature of both likelihood functions and the role, the ω_j and ω_j^* play in them, the nonzero components of the log likelihood's maximum are, as well, the solutions of maximizing the corresponding reduced version of the log likelihood function.

Due to this equality, whose proof is shown in Section B.3, we assume that the proposed reduced version of the likelihood will act as the usual likelihood, in the sense that $(\hat{\boldsymbol{\theta}}_n^*, \hat{\boldsymbol{\omega}}_n^*)$ is asymptotically normal and we can use the associated Fisher information to determine the confidence intervals of μ, β, and σ. We carry out the computation of the first and second derivative of $l^*(\boldsymbol{\theta}, \boldsymbol{\omega}^*) = \ln\big(L^*(\boldsymbol{\theta}, \boldsymbol{\omega}^*)\big)$ with the use of the mathematical software package Maple. Then, the inverse of the observed information matrix serves as an estimation of the asymptotic covariance matrix (Collett 1994, Appendix A).

3.6 Further aspects

The estimation procedure described above can also be applied in the following two cases: further covariates are included in the model and observations with missing values in the response are considered.

3.6.1 Model extensions

The accelerated failure time model (3.5) can easily be extended to include other covariates summarized in the vector $\boldsymbol{X}$:

$$\ln(Y) = \mu + \beta Z + \boldsymbol{\kappa' X} + \sigma W.$$

This affects the density and survival functions of Y in (3.24) and (3.29) as follows:

$$\begin{aligned} S_Y(y) &= S_W\big(\tfrac{\ln(y)-\mu-\beta z-\boldsymbol{\kappa' x}}{\sigma}\big), \\ f_Y(y) &= \tfrac{1}{\sigma} f_W\big(\tfrac{\ln(y)-\mu-\beta z-\boldsymbol{\kappa' x}}{\sigma}\big). \end{aligned} \tag{3.34}$$

If the covariates summarized in $\boldsymbol{X}$ are observed completely, S_W and f_W in (3.34) are plugged into the (log) likelihood functions above using the corresponding expressions for the Weibull and the log logistic regression models.

If one of the covariates, V say, is interval-censored with observed intervals $[V_L, V_R]$, such that $\mathrm{P}(V \in [V_L, V_R]) = 1$, the likelihood contributions have to consider all possible values of

V given $[V_L, V_R]$. We illustrate this for the case of a left- and right-censored response variable assuming noninformative censoring following the conditions on page 44. Application to the case of a doubly-censored response variable is straightforward.

Given the observation vector $(u, z_l, z_r, v_l, v_r, \delta_1, \delta_2)$ for a single individual, its likelihood contribution is an extension of expression (3.12) on page 45:

$$L_{cont}(\boldsymbol{\theta}, F_Z, F_V) = \int_{z_l}^{z_r} \int_{v_l}^{v_r} f(u|z,v)^{\delta_1} S(u|z,v)^{\delta_2} \big(1 - S(u|z,v)\big)^{(1-\delta_1)(1-\delta_2)} \, dF_V(v) \, dF_Z(z),$$

where F_V denotes the distribution function of V. The resulting likelihood function given independent observation vectors is the product over all likelihood contributions. A discrete version of the likelihood function is obtained by proceeding as for case 1 in Section 3.3.2 for both Z and V.

3.6.2 Inclusion of missing values of the response variable

Missing values of the response variable Y might be present. For example, if Y is a subsequent time of the covariate Z, a right-censored value of the latter implies a missing value of the former. Given the structure of the likelihood functions above, these observations can still be included. They contribute to the estimation of F_Z, equivalent to $\boldsymbol{\omega}$ in the discrete case, and, for this reason, they might improve the estimation of the unknown model parameters. The likelihood contribution of an individual with missing value in the response variable is $\int_{z_l}^{z_r} dF_Z(z)$. To incorporate this term in the (log) likelihood functions, another indicator function must be defined:

$$\epsilon = \begin{cases} 1 & Y \text{ is exactly observed or censored} \\ 0 & Y \text{ is missing} \end{cases} .$$

When $\epsilon = 1$, the value of the response variable might be either exactly observed or left- or right-censored, whereas if Y is missing, $\epsilon = \delta_1 = \delta_2 = 0$. Including this indicator variable, the observation vectors (3.8) and (3.9) are extended to $(U, Z_l, Z_r, \epsilon, \delta_1, \delta_2)$ and $(U, Y_{0l}, Y_{0r}, Z_l, Z_r, \epsilon, \delta_1, \delta_2)$, respectively.

For example, consider the likelihood functions for the case of a left- and right-censored response variable, when the covariate is either continuous or discrete. Allowing for missing values in Y, their expressions are the following:

$$L(\boldsymbol{\theta}, F_Z) = \prod_{i=1}^{n} \int_{z_{l_i}}^{z_{r_i}} f(u_i|z)^{\epsilon_i \delta_{1_i}} S(u_i|z)^{\epsilon_i \delta_{2_i}} \big(1 - S(u_i|z)\big)^{\epsilon_i (1-\delta_{1_i})(1-\delta_{2_i})} \, dF_Z(z),$$

$$L(\boldsymbol{\theta}, \boldsymbol{\omega}) = \prod_{i=1}^{n} \sum_{j=1}^{m} \alpha_{ij} f(u_i|s_j)^{\epsilon_i \delta_{1_i}} S(u_i|s_j)^{\epsilon_i \delta_{2_i}} \big(1 - S(u_i|s_j)\big)^{\epsilon_i (1-\delta_{1_i})(1-\delta_{2_i})} \omega_j. \tag{3.35}$$

Note that in the discrete case (3.35), the likelihood contribution of an individual with $\epsilon_i = 0$ is equal to $\sum_{j=1}^{m} \alpha_{ij}\omega_j$.

Hence, the proposed methodology for the case an accelerated failure time model with an interval-censored covariate allows for the inclusion of observations with missing response in the likelihood. However, the interpretation of the estimation results of the reduced data evaluation needs to consider the randomness of the missing data generation process.