Chapter 4

Simultaneous Maximization of the Likelihood of Further Regression Models with an Interval-Censored Covariate

The accelerated failure time model of the previous chapter has been motivated by the data set on injecting drug users in Badalona, whose complex censoring patterns in both the response and the covariate has made it necessary to develop alternative estimation procedures. Such particular data patterns are not restricted to survival analysis only, but might arise in any other area. Besides censoring, think of longitudinal data, for example.

Generally, when applying a regression model to such data, the resulting likelihood function does not only depend on the model parameters, but also on nuisance parameters. As shown in Chapter 3, in case of an interval-censored covariate, these parameters account for the unknown distribution function of that covariate. To the best of our knowledge, various types of EM algorithms or imputation techniques have been implemented to carry out parameter estimation in the presence of such nuisance parameters.

In this chapter, we show that the mathematical programming language AMPL in combination with the NEOS solvers can be a very valuable tool for the evaluation of regression models, whose evaluation requires the implementation of estimation procedures not covered by statistical software. AMPL permits the maximization of likelihood functions without the need to programme iterative estimation procedures. In Section 4.1, we describe the general steps of the estimation procedure. After this, we illustrate the utility of AMPL for three other models with an interval-censored covariate: linear regression, logistic regression, and the semi-parametric Cox model. For each of these, we derive the corresponding likelihood function. In Section C.1, parts of the corresponding AMPL programmes to compute the parameter estimates are provided.

4.1 Description of the general estimation procedure

The estimation procedure described herein consists of the following five steps and is valid for any regression model, for which one is able to derive the likelihood function. Obviously, whenever statistical software is available for the evaluation of a given model, its use is preferable to the estimation procedure with AMPL.

1. Derivation of the likelihood function

   The main theoretical work consists in deriving the correct likelihood taking into account possible dependencies or censoring patterns of the data. In Chapter 3 and in the following examples, noninformative censoring is assumed. The estimation procedure does not require that assumption, but otherwise the censoring mechanism has to be modeled, which implies further assumptions on the data generation process and hence more nuisance parameters.

2. Implementation of the maximization of the likelihood by means of AMPL

   The programming code of AMPL is quite intuitive which helps to programme the maximization of the likelihood function with respect to all parameters and subject to possible restrictions. Generally, it is easier to carry out the maximization of the log likelihood. The AMPL code is presented in detail in Fourer, Gay, and Kernighan (2003) and illustrated in Section C.1 for each of the following examples.

   If the number of nuisance parameters is infinite because of the presence of a parametrically undefined continuous function, assumptions have to be made concerning its form, for example, piecewise linear or constant, or the function must be assumed discrete on a defined grid of points.

3. Execution of the AMPL programme invoking an adequate NEOS solver

   Often, the available solvers of the local version of AMPL are not able to solve the maximization problem. For this reason, the NEOS server offers the use of several solvers for plenty of optimization problems written with the AMPL code; the use of these solvers is free of charge (Dolan, Fourer, Moré, and Munson 2002).

   Dealing with maximum likelihood estimation, the objective function is mostly nonlinear and maximization is subject to several restrictions. Possible solvers for these nonlinearly constrained optimization problems are MINOS (Murtagh and Saunders 1978), LANCELOT (Conn, Gould, and Toint 1992), or SNOPT (Gill, Murray, and Saunders 1999). The latter has been proved to be an adequate solver for the maximization problem presented in this and the previous chapter. Some details on NEOS and SNOPT are described on page 14.

4. Collection of estimation results

   If one of the NEOS solvers is used during the execution of the AMPL programmes, the results can either be received by email or can be written directly into a file on the local disc. In both cases, the results have to be edited in such a way, that the software used in the final step can read the data.

5. Computation of confidence intervals using Maple

   Any mathematical software can be used for this final step, Maple being one of these. The reduced log likelihood —see Section 3.5.3— must be derived twice, the observed Fisher matrix is computed, and finally inverted. The resulting matrix is the covariance matrix, whose diagonal elements are used for the computation of the confidence intervals.

4.2 Linear regression model

A linear regression model with a discrete interval-censored covariate is presented by Gómez, Espinal, and Lagakos (2003). The data, that have motivated the use of such a model, come from the study 359 of the AIDS Clinical Trial Group on HIV patients who have previously failed a treatment with an antiviral therapy. The interest of the authors has focused on the possible relation between the waiting time from treatment failure until study enrolment, the (interval-censored) covariate, and the viral load level at the time of enrolment, the response variable.

4.2.1 Model and likelihood function

The model considered here is similar to the one of Gómez et al. presented before in Section 1.1.4 on page 9. Here, we consider the same model including also the fully observed covariate vector $\boldsymbol{X}$:

$$Y = \alpha + \beta Z + \boldsymbol{\kappa}'\boldsymbol{X} + \epsilon, \tag{4.1}$$

where $\epsilon \sim \mathcal{N}(0, \sigma^2)$ and independent of $\boldsymbol{X}$ and Z, the discrete interval-censored covariate. The support of Z is given by $S = \{s_1, \ldots, s_m\}$ with probabilities $\mathrm{P}(Z = s_j) = \omega_j$, $j = 1, \ldots, m$, and the observed intervals are denoted by $[Z_l, Z_r]$. The objective consists in estimating the unknown parameter vector $\boldsymbol{\theta} = (\alpha, \beta, \boldsymbol{\kappa}, \sigma)'$ given the independent observations $(y_i, z_{l_i}, z_{r_i}, \boldsymbol{x}_i)$, $i = 1, \ldots, n$, where $\mathrm{P}(z_i \in [z_{l_i}, z_{r_i}]) = 1$.

Under the assumption that censoring is noninformative and defining the indicator variables $\gamma_{ij} = \mathbb{1}_{\{s_j \in [z_{l_i}, z_{r_i}]\}}$, the likelihood function of model (4.1) is proportional to

$$\begin{aligned} L(\boldsymbol{\theta}, \boldsymbol{\omega}) &= \prod_{i=1}^{n} \sum_{j=1}^{m} \gamma_{ij} f(y_i | s_j, \boldsymbol{x}_i; \boldsymbol{\theta}) \omega_j \\ &= \prod_{i=1}^{n} \sum_{j=1}^{m} \gamma_{ij} \frac{1}{\sqrt{2\pi}\sigma} \exp\Big(- \frac{(y_i - \alpha - \beta s_j - \boldsymbol{\kappa}'\boldsymbol{x}_i)^2}{2\sigma^2} \Big) \omega_j, \end{aligned}$$

where $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_m)'$. The noninformativity conditions concerning censoring in Z hold, if the following equalities are valid for any values y, z, z_l, and z_r:

$$\begin{aligned} \mathrm{P}(Z = z | Z_l = z_l, Z_r = z_r) &= \mathrm{P}(Z = z)/\mathrm{P}(z_l \leq Z \leq z_r) \mathbb{1}_{\{z \in [z_l, z_r]\}}, \\ f(y | Z = z, Z_l = z_l, Z_r = z_r) &= f(y | Z = z). \end{aligned} \tag{4.2}$$

4.2.2 Algorithms to maximize the likelihood

The estimation procedure proposed by Gómez et al. for the computation of the joint maximum likelihood estimator $(\hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\omega}}_n)$ is described in Section 1.1.4. Their algorithm, programmed in C, alternates two steps —estimation of $\boldsymbol{\omega}$ holding $\boldsymbol{\theta}$ fixed and viceversa— until simultaneous convergence is achieved.

The same algorithm could be programmed with the AMPL code alternating both estimation steps by means of the `problem` statement (Fourer et al., Chap. 14.4). However, disposing of the NEOS solvers in AMPL, it is easier and computationally more efficient to carry out simultaneous maximization of the likelihood function with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\omega}$. For computational reasons it is preferable to use the log likelihood as objective function:

$$l(\boldsymbol{\theta}, \boldsymbol{\omega}) = \sum_{i=1}^{n} \ln \Big(\sum_{j=1}^{m} \gamma_{ij} \exp \big(- \frac{(y_i - \alpha - \beta s_j - \boldsymbol{\kappa}' \boldsymbol{x}_i)^2}{2\sigma^2} \big) \omega_j \Big) - n \ln(\sqrt{2\pi}\sigma). \tag{4.3}$$

The maximization of (4.3) is subject to the constraints on $\boldsymbol{\omega}$:

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

$$\omega_j \geq 0, \; j = 1, \ldots, m.$$

Once the estimators $\hat{\boldsymbol{\theta}}_n$ and $\hat{\boldsymbol{\omega}}_n$ are computed, their variances can be estimated in the same way as described in Section 3.5.3 for the accelerated failure time model, using the mathematical software Maple.

4.2.3 Model extensions

So far, the response variable has been assumed to be completely observed. Nonetheless, Y might also be censored. Consider the case of right- and left-censoring at time y_i. In these cases, the likelihood contributions of an individual i are equal to

$$C_i(\boldsymbol{\theta}, \boldsymbol{\omega}) = \sum_{j=1}^{m} \gamma_{ij} S(y_i | s_j, \boldsymbol{x}_i) \omega_j = \sum_{j=1}^{m} \gamma_{ij} \big(1 - \Phi(y_i | s_j, \boldsymbol{x}_i)\big) \omega_j$$

and

$$C_i(\boldsymbol{\theta}, \boldsymbol{\omega}) = \sum_{j=1}^{m} \gamma_{ij}\big(1 - S(y_i|s_j, \boldsymbol{x}_i)\big)\omega_j = \sum_{j=1}^{m} \gamma_{ij}\Phi(y_i|s_j, \boldsymbol{x}_i)\omega_j,$$

respectively, being Φ the Gaussian distribution function. Analogously to the likelihood function of the accelerated failure time model in Section 3.3, indicator functions δ_1 and δ_2 must be defined to distinguish the different censoring patterns:

$$\delta_1 = \begin{cases} 1 & Y \text{ observed exactly} \\ 0 & \text{otherwise} \end{cases},$$

$$\delta_2 = \begin{cases} 1 & Y \text{ right-censored} \\ 0 & \text{otherwise} \end{cases}.$$

Then, given the independent observations $(y_i, z_{l_i}, z_{r_i}, \boldsymbol{x}_i, \delta_{1_i}, \delta_{2_i})$, $i = 1, \ldots, n$, the resulting likelihood function is the following:

$$L(\boldsymbol{\theta}, \boldsymbol{\omega}) =$$
$$\prod_{i=1}^{n}\sum_{j=1}^{m} \gamma_{ij}\Big(\frac{1}{\sqrt{2\pi}\sigma}\exp\big(-\frac{(y_i - \alpha - \beta s_j - \boldsymbol{\kappa}'\boldsymbol{x}_i)^2}{2\sigma^2}\big)\Big)^{\delta_{1_i}} \big(1 - \Phi(y_i|s_j, \boldsymbol{x}_i)\big)^{\delta_{2_i}} \Phi(y_i|s_j, \boldsymbol{x}_i)^{(1-\delta_{1_i})(1-\delta_{2_i})}\omega_j.$$

The inclusion of interval-censored data $[y_{l_i}, y_{r_i}]$ is straightforward with likelihood contributions equal to $\sum_{j=1}^{m} \gamma_{ij}\big(\Phi(y_{r_i}|s_j, \boldsymbol{x}_i) - \Phi(y_{l_i}|s_j, \boldsymbol{x}_i)\big)\omega_j$, but requires one more indicator variable.

Apart from the normal distribution, Y might follow any other distribution. This would affect the expressions of the density and survival functions, but not the general form of the likelihood functions.

4.3 Logistic regression with an interval-censored covariate

4.3.1 The logistic regression model

Logistic regression models the probability of the occurrence of an event of interest in dependence of one or more covariates. For example, in biostatistics, this event of interest is often the presence of a disease or death. We denote this dichotomized random variable by Y:

$$Y = \begin{cases} 1 & \text{Event occurred} \\ 0 & \text{Otherwise} \end{cases}.$$

Consider the case that one covariate, Z say, is censored in an interval: $Z \in [Z_l, Z_r]$. Summarizing the other covariates of the model in the vector $\boldsymbol{X}$, the logistic regression model can be expressed as follows:

$$p = \mathrm{P}(Y = 1|Z, \boldsymbol{X}) = \frac{\exp(\alpha + \beta Z + \boldsymbol{\kappa}'\boldsymbol{X})}{1 + \exp(\alpha + \beta Z + \boldsymbol{\kappa}'\boldsymbol{X})},$$

which is equivalent to

$$\ln\Big(\frac{p}{1-p}\Big) = \alpha + \beta Z + \boldsymbol{\kappa}'\boldsymbol{X}.$$

As before, we assume that Z is discrete with possible values $s_1 < s_2 < \cdots < s_m$ and corresponding probabilities $\omega_j = \mathrm{P}(Z = s_j)$, $j = 1, \ldots, m$. The objective consists in estimating the unknown parameter vector $\boldsymbol{\theta} = (\alpha, \beta, \boldsymbol{\kappa})'$ by means of maximum likelihood estimation.

4.3.2 Likelihood functions

As with the models before, we assume that censoring in the covariate Z is noninformative, that is, the observed interval $[Z_l, Z_r]$ does not inform on the real value of Z apart from including it, and it has no influence on the response variable Y:

$$\mathrm{P}(Z = z|Z_l = z_l, Z_r = z_r) = \mathrm{P}(Z = z)/\mathrm{P}(z_l \leq Z \leq z_r)\mathbb{1}_{\{z \in [z_l, z_r]\}},$$
$$\mathrm{P}(Y = 1|Z = z, Z_l = z_l, Z_r = z_r) = \mathrm{P}(Y = 1|Z = z).$$

If Z was observed exactly, the likelihood contribution of an individual with observed values $(y, z, \boldsymbol{x})$ would be equal to:

$$\begin{aligned} C(\boldsymbol{\theta}) = p^y(1-p)^{1-y} &= \Big(\frac{\exp(\alpha + \beta z + \boldsymbol{\kappa}'\boldsymbol{x})}{1 + \exp(\alpha + \beta z + \boldsymbol{\kappa}'\boldsymbol{x})}\Big)^y \Big(\frac{1}{1 + \exp(\alpha + \beta z + \boldsymbol{\kappa}'\boldsymbol{x})}\Big)^{1-y} \\ &= \frac{\exp(\alpha + \beta z + \boldsymbol{\kappa}'\boldsymbol{x})^y}{1 + \exp(\alpha + \beta z + \boldsymbol{\kappa}'\boldsymbol{x})}. \end{aligned}$$

However, since Z lies in the interval $[Z_l, Z_r]$, the likelihood contribution has to account for all admissible values of Z given that interval. That is, the likelihood contribution of an individual i depends also on the distribution function of Z characterized by the vector $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_m)'$:

$$C_i(\boldsymbol{\theta}, \boldsymbol{\omega}) = \sum_{j=1}^{m} \gamma_{ij} \frac{\exp(\alpha + \beta s_j + \boldsymbol{\kappa}'\boldsymbol{x}_i)^y}{1 + \exp(\alpha + \beta s_j + \boldsymbol{\kappa}'\boldsymbol{x}_i)} \omega_j,$$

where the binary variables $\gamma_{ij} = \mathbb{1}_{\{s_j \in [z_{l_i}, z_{r_i}]\}}$ indicate whether s_j is an admissible value for z_i or not. Consequently and supposing that the observations $(y_i, z_{l_i}, z_{r_i}, \boldsymbol{x}_i)$, $i = 1, \ldots, n$, are independent, the likelihood function for the unknown model parameters is proportional to

$$\begin{aligned}
L(\boldsymbol{\theta},\boldsymbol{\omega}) &= \prod_{i=1}^{n} C_i(\boldsymbol{\theta},\boldsymbol{\omega}) \\
&= \prod_{i=1}^{n}\sum_{j=1}^{m} \gamma_{ij} \frac{\exp(\alpha+\beta s_j+\boldsymbol{\kappa}'\boldsymbol{x}_i)^{y_i}}{1+\exp(\alpha+\beta s_j+\boldsymbol{\kappa}'\boldsymbol{x}_i)}\omega_j \\
&= \prod_{i=1}^{n} \exp\left(y_i(\alpha+\boldsymbol{\kappa}'\boldsymbol{x}_i)\right) \sum_{j=1}^{m} \gamma_{ij} \frac{\exp(y_i\beta s_j)}{1+\exp(\alpha+\beta s_j+\boldsymbol{\kappa}'\boldsymbol{x}_i)}\omega_j.
\end{aligned}$$

Maximizing $L(\boldsymbol{\theta},\boldsymbol{\omega})$ furnishes the maximum likelihood estimators of the model parameters α, β, and $\boldsymbol{\kappa}$, as well as $\hat{\boldsymbol{\omega}}_n$. Instead of maximizing the likelihood function, the estimation is accomplished more easily by maximizing the log likelihood function:

$$l(\boldsymbol{\theta},\boldsymbol{\omega}) = \sum_{i=1}^{n}\left(y_i(\alpha+\boldsymbol{\kappa}'\boldsymbol{x}_i) + \ln\left(\sum_{j=1}^{m}\gamma_{ij}\frac{\exp(y_i\beta s_j)}{1+\exp(\alpha+\beta s_j+\boldsymbol{\kappa}'\boldsymbol{x}_i)}\omega_j\right)\right). \tag{4.4}$$

Hence, the optimization problem for the estimation of $\boldsymbol{\theta}$ and $\boldsymbol{\omega}$ consists of the objective function (4.4) and the restrictions on $\boldsymbol{\omega}$:

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

$$\omega_j \geq 0,\ j = 1,\ldots,m.$$

4.4 The proportional hazards model

The proportional hazards model (Cox 1972) is well known and popular for its semi-parametric structure, which allows to separate the estimation of the covariates' effect from the estimation of the baseline hazard function. Its general form for a survival time T and covariate vector $\boldsymbol{Z}$, which models the hazard function λ, is given by

$$\lambda(t;\boldsymbol{z}) = \lambda_0(t)\exp(\boldsymbol{\beta}'\boldsymbol{z}), \tag{4.5}$$

where $\lambda_0(t)$ is the baseline hazard function and $\boldsymbol{\beta}$ the unknown parameter vector of interest. The particularity we are dealing with, is the case of an interval-censored covariate Z. For the sake of simplicity, we consider the case of model (4.5) with only one covariate and our aim is to estimate the unknown parameter β.

4.4.1 Derivation of the partial likelihood function

Due to its semi-parametric structure leaving the baseline hazard function unspecified, the likelihood function of the Cox model does not only depend on β, but also on λ_0. However, for the parameter estimation, it is possible to use the partial likelihood function, in which λ_0 does not appear. We, therefore, first recall the equality of Cox's partial likelihood function with the profile likelihood for the case of a fully observed covariate Z. After this, we attempt to apply the same procedure to the case of an interval-censored covariate.

The likelihood functions are derived given independent observations (t_i, z_i, δ_i), $i = 1, \dots, n$, where $\delta_i = 1$ indicates an exactly observed survival time t_i, and $\delta_i = 0$ a right-censored one. For the present, we do not consider other censoring patterns rather than right-censoring and assume noninformative censoring in T.

Cox proposes the use of the partial likelihood function for the estimation of the parameter β in model (4.5). Assuming there are no ties among the uncensored survival times, this function has the following expression:

$$L(\beta) = \prod_{i=1}^{n} \Big(\frac{\exp(\beta z_i)}{\sum_{j \in R(t_i)} \exp(\beta z_j)} \Big)^{\delta_i}, \tag{4.6}$$

where $R(t_i)$ is the group of subjects at risk just before t_i. This function does not depend on the unknown baseline hazard function and is equivalent to the profile likelihood function for β, $L_{prof}(\beta)$ (Murphy and van der Vaart 2000).

To show this equivalence, consider the full likelihood function under noninformative censoring including both β and the baseline hazard function λ_0:

$$\begin{aligned} L\big(\beta, \lambda_0(\cdot)\big) &= \prod_{i=1}^{n} f(t_i)^{\delta_i} S(t_i)^{1-\delta_i} = \prod_{i=1}^{n} \big(\lambda(t_i) S(t_i)\big)^{\delta_i} S(t_i)^{1-\delta_i} = \prod_{i=1}^{n} \lambda(t_i)^{\delta_i} S(t_i) \\ &= \prod_{i=1}^{n} \big(\lambda_0(t_i) \exp(\beta z_i)\big)^{\delta_i} \exp\big(- \Lambda_0(t_i) \exp(\beta z_i)\big). \end{aligned} \tag{4.7}$$

Hence, the log likelihood has the following expression:

$$l\big(\beta, \lambda_0(\cdot)\big) = \sum_{i=1}^{n} \Big(\delta_i \big(\ln(\lambda_0(t_i)) + \beta z_i \big) - \Lambda_0(t_i) \exp(\beta z_i) \Big). \tag{4.8}$$

where $\Lambda_0(t_i)$ is the baseline cumulative hazard function. The profile likelihood function for β is defined by:

$$L_{prof}(\beta) = \sup_{\lambda_0} L(\beta, \lambda_0).$$

Hence, to determine the profile likelihood function, the likelihood function (4.7) has to be maximized with respect to $\lambda_0(t)$ holding β fixed. Without any smoothing constraints, $\hat{\Lambda}_0$ will be a step function with jumps at t_i, $i = 1, \dots, n$, whenever $\delta_i = 1$. The jumps are equal to $\hat{\lambda}_0(t_i)$. Deriving the log likelihood (4.8) with respect to $\lambda_0(t_i)$, we obtain

$$\frac{\partial}{\partial \lambda_0(t_i)} l\big(\beta, \lambda_0(\cdot)\big) = \frac{\delta_i}{\lambda_0(t_i)} - \sum_{j:t_j \geq t_i} \exp(\beta z_j).$$

Thus, the likelihood is maximized for

$$\hat{\lambda}_0(t_i) = \frac{\delta_i}{\sum_{j:t_j \geq t_i} \exp(\beta z_j)}, \; i = 1, \dots, n. \tag{4.9}$$

Note, that $\hat{\lambda}_0(t) = 0$ if $\delta = 0$ as well as for any $t \notin \{t_1, \dots, t_n\}$. Plugging (4.9) into (4.7), we obtain the partial likelihood function (4.6):

$$\begin{aligned}
L(\beta) &= \prod_{i=1}^{n} \big(\hat{\lambda}_0(t_i) \exp(\beta z_i)\big)^{\delta_i} \exp\big(- \hat{\Lambda}_0(t_i) \exp(\beta z_i)\big) \\
&= \prod_{i=1}^{n} \Big(\frac{\exp(\beta z_i)}{\sum_{j:t_j \geq t_i} \exp(\beta z_j)} \Big)^{\delta_i} \exp\Big(- \exp(\beta z_i) \sum_{j:t_j \leq t_i} \hat{\lambda}_0(t_j) \Big) \\
&= \prod_{i=1}^{n} \Big(\frac{\exp(\beta z_i)}{\sum_{j \in R(t_i)} \exp(\beta z_j)} \Big)^{\delta_i} \exp\Big(- \exp(\beta z_i) \sum_{j:t_j \leq t_i} \frac{\delta_j}{\sum_{l:t_l \geq t_j} \exp(\beta z_l)} \Big) \\
&= \exp\Big(- \sum_{i=1}^{n} \exp(\beta z_i) \sum_{j:t_j \leq t_i} \frac{\delta_j}{\sum_{l:t_l \geq t_j} \exp(\beta z_l)} \Big) \prod_{i=1}^{n} \Big(\frac{\exp(\beta z_i)}{\sum_{j \in R(t_i)} \exp(\beta z_j)} \Big)^{\delta_i} \\
&= \exp\Big(- \sum_{j=1}^{n} \sum_{i:t_i \geq t_j} \frac{\delta_j \exp(\beta z_i)}{\sum_{l:t_l \geq t_j} \exp(\beta z_l)} \Big) \prod_{i=1}^{n} \Big(\frac{\exp(\beta z_i)}{\sum_{j \in R(t_i)} \exp(\beta z_j)} \Big)^{\delta_i} \\
&= \exp\Big(- \sum_{j=1}^{n} \delta_j \frac{\sum_{i:t_i \geq t_j} \exp(\beta z_i)}{\sum_{l:t_l \geq t_j} \exp(\beta z_l)} \Big) \prod_{i=1}^{n} \Big(\frac{\exp(\beta z_i)}{\sum_{j \in R(t_i)} \exp(\beta z_j)} \Big)^{\delta_i} \\
&= \prod_{i=1}^{n} \Big(\frac{\exp(\beta z_i)}{\sum_{j \in R(t_i)} \exp(\beta z_j)} \Big)^{\delta_i} \exp(-\delta_i) \\
&\propto \prod_{i=1}^{n} \Big(\frac{\exp(\beta z_i)}{\sum_{j \in R(t_i)} \exp(\beta z_j)} \Big)^{\delta_i}.
\end{aligned}$$

4.4.2 Likelihood in presence of an interval-censored covariate

Now, the question is, how to handle the case of an interval-censored covariate? That is, instead of Z, we observe the interval $[Z_L, Z_R]$ containing Z. Concerning the censoring generation process, we assume the noninformativity conditions (4.2) of the previous sections hold.

Departing from likelihood function (4.7), we derive the following likelihood, which takes into account the intervals $[Z_{L_i}, Z_{R_i}],\ i = 1, \ldots, n$. The distribution function of Z is denoted by F_Z, its density function by f_Z.

$$L_n\big(\beta, \lambda_0(\cdot), F_Z(\cdot)\big) = \prod_{i=1}^{n} \int_{z_{l_i}}^{z_{r_i}} \big(\lambda_0(t_i)\exp(\beta z)\big)^{\delta_i} \exp\big(-\Lambda_0(t_i)\exp(\beta z)\big)\, dF_Z(z)$$
$$= \prod_{i=1}^{n} \int_{z_{l_i}}^{z_{r_i}} \big(\lambda_0(t_i)\exp(\beta z)\big)^{\delta_i} \exp\big(-\Lambda_0(t_i)\exp(\beta z)\big) f_Z(z)\, dz.$$

In case of a discrete covariate Z with support $S = \{s_1, \ldots, s_m\}$, where $s_1 < s_2 < \cdots < s_m$, and corresponding probabilities $P(Z = s_j) = \omega_j,\ j = 1, \ldots, m$, the preceding likelihood function can be written as

$$L_n\big(\beta, \lambda_0(\cdot), \boldsymbol{\omega}\big) = \prod_{i=1}^{n}\sum_{j=1}^{m} \alpha_{ij}\big(\lambda_0(t_i)\exp(\beta s_j)\big)^{\delta_i} \exp\big(-\Lambda_0(t_i)\exp(\beta s_j)\big)\omega_j, \qquad (4.10)$$

in which $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_m)'$ and $\alpha_{ij} = \mathbb{1}_{\{s_j \in (z_{l_i}, z_{r_i}]\}}$.

4.4.3 Maximization procedures

Attempt by means of the profile likelihood

Following the procedure in Section 4.4.1, an attempt is made to obtain the maximum likelihood estimators by means of the profile likelihood in order to simplify the expression of the likelihood function (4.10) to be maximized. This is done for the case of a discrete covariate. Consider the log likelihood:

$$l\big(\beta, \lambda_0(\cdot), \boldsymbol{\omega}\big) = \sum_{i=1}^{n}\Bigg(\delta_i \ln\big(\lambda_0(t_i)\big) + \ln\Bigg(\sum_{j=1}^{m} \alpha_{ij}\exp(\delta_i\beta s_j)\exp\big(-\Lambda_0(t_i)\exp(\beta s_j)\big)\omega_j\Bigg)\Bigg). \qquad (4.11)$$

Denoting $\lambda_0(t_i)$ by λ_i, the differentiation of (4.11) furnishes

$$\frac{\partial}{\partial\lambda_k} l_n\big(\beta, \lambda_0(\cdot), \boldsymbol{\omega}\big) = \frac{\delta_k}{\lambda_k} + \sum_{i:t_i \geq t_k}^{n} \frac{\lambda_k \sum_{j=1}^{m} \alpha_{ij}\exp\big((1+\delta_i)\beta s_j\big)\exp\big(-\sum_{r:t_r\leq t_i}\lambda_r\exp(\beta s_j)\big)\omega_j}{\sum_{j=1}^{m}\alpha_{ij}\exp(\delta_i\beta s_j)\exp\big(-\sum_{r:t_r\leq t_i}\lambda_r\exp(\beta s_j)\big)\omega_j}. \qquad (4.12)$$

Since λ_k cannot be factorized in the second summand of expression (4.12), there is no closed form of $\hat{\lambda}_k$. Consequently, the estimates $\hat{\lambda}_k,\ k = 1, \ldots, n$, would have to be computed by means of numerical methods whenever $\delta_k = 1$. The obtained values would then be plugged into (4.11) and maximization procedures would have to be applied in order to calculate $\hat{\beta}$ and $\hat{\boldsymbol{\omega}}$. In case of a continuous covariate, the whole procedure would be even more cumbersome, since the integrals would have to be approximated.

However, since the motivation of the profile likelihood approach is to facilitate the maximization procedure, there is no evident advantage of this method here. As commented below, simultaneous maximization of the log likelihood function (4.11) is preferable.

If $\boldsymbol{\omega}$ was estimated in a first step, for example applying the Turnbull estimator of page 2, the profile likelihood approach would have to tackle the same problem. Also, profiling out $\boldsymbol{\omega}$ in the log likelihood (4.11) does not furnish a closed form for $\hat{\boldsymbol{\omega}}$, neither. The reason can be seen in the following differentiation of (4.11) with respect to any ω_k:

$$\frac{\partial}{\partial \omega_k} l_n\big(\beta, \lambda_0(\cdot), \boldsymbol{\omega}\big) = \sum_{i=1}^{n} \frac{\alpha_{ik} \exp(\delta_i \beta s_k) \exp\big(-\Lambda_0(t_i)\exp(\beta s_j)\big)}{\sum_{j=1}^{m} \alpha_{ij} \exp(\delta_i \beta s_j) \exp\big(-\Lambda_0(t_i)\exp(\beta s_j)\big)\omega_j}.$$

Simultaneous parameter estimation: full likelihood approach

Taking advantage of the mathematical programming language AMPL and its facility to invoke the NEOS solvers, we propose the simultaneous maximization of the log likelihood (4.11) with respect to β, λ_0, and $\boldsymbol{\omega}$. This can be done under the assumption that the response variable T is discrete with support equal to the uncensored times t_i. This nonlinear maximization problem is subject to the equality constraint

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

and the inequality constraints

$$\begin{aligned} \omega_j &\geq 0,\ j = 1, \ldots, m, \\ \lambda_0(t) &\geq 0,\ t > 0. \end{aligned} \tag{4.13}$$

To accomplish the maximization procedure, in the objective function (4.11) we either

- substitute $\Lambda_0(t_i)$ by $\sum_{j: t_j \leq t_i} \lambda_0(t_j)$ and estimate $\lambda_0(t_i)$, $i = 1, \ldots, n$, or
- substitute $\lambda_0(t_i)$ by $\Lambda_0(t_i) - \Lambda_0(t_{i-1})$ and estimate $\Lambda_0(t_i)$, $i = 1, \ldots, n$.

Note that the latter requires ordering the survival times: $t_1 \leq t_2 \leq \cdots \leq t_n$ and that the inequality constraint (4.13) implies $0 \leq \Lambda_0(t_1) \leq \Lambda_0(t_2) \leq \cdots \leq \Lambda_0(t_n)$. Simulations will have to show, whether this procedure is feasible or not.

4.4.4 A pseudo likelihood approach

Another approach for the estimation of β in model (4.5), is the use of pseudo likelihood functions (Gouriéroux and Monfort 1993), which are not derived from the full likelihood (4.7). For example, one could think of departing from the partial likelihood (4.6) taking into account all values of the intervals $[Z_{l_i}, Z_{r_i}]$, $i = 1, \dots, n$:

$$L(\beta) = \prod_{i=1}^{n} \sum_{j=1}^{m} \alpha_{ij} \Big(\frac{\exp(\beta s_j)}{\sum_{j \in R(t_i)} \exp(\beta z_j)} \Big)^{\delta_i} \omega_j.$$

However, this function neglects that in the denominator all combinations of values of the subjects of the risk set would have to be considered. That is, for a given i, all combinations of values $z \in [z_{l_k}, z_{r_k}]$ with $k \geq i$ would have to be considered. This is infeasible, even for small sample sizes.

In contrast with that, the unobserved values z_i in the log likelihood function of model (4.5),

$$l\big(\beta, \Lambda_0(\cdot)\big) = \sum_{i=1}^{n} \Big(\delta_i \big(\ln(\lambda_0(t_i)) + \beta z_i \big) - \Lambda_0(t_i) \exp(\beta z_i) \Big),$$

could be replaced by its expected value conditioned on the observed interval:

$$\mathrm{E}_{F_Z}(Z | Z_L, Z_R) = \int_{Z_L}^{Z_R} Z \, dF_Z = \int_{z_l}^{z_r} z f_Z(z) \, dz.$$

In case of a discrete Z, this expected value is equal to $\sum_{j=1}^{m} \alpha_{ij} s_j \omega_j$. Plugging that expression into the log likelihood, we obtain the pseudo log likelihood function

$$l_{ps}\big(\beta, \lambda_0(\cdot), \boldsymbol{\omega}\big) = \sum_{i=1}^{n} \Big(\delta_i \big(\ln(\lambda_0(t_i)) + \beta \sum_{j=1}^{m} \alpha_{ij} s_j \omega_j \big) - \Lambda_0(t_i) \exp\big(\beta \sum_{j=1}^{m} \alpha_{ij} s_j \omega_j \big) \Big).$$

For this function, the same procedure as in Section 4.4.1 can be applied: profiling out λ_0, furnishes the following function, the so-called partial pseudo likelihood function:

$$L_{ps}(\beta) = \prod_{i=1}^{n} \Big(\frac{\exp(\beta \sum_{j=1}^{m} \alpha_{ij} s_j \omega_j)}{\sum_{l \in R(t_i)} \exp(\beta \sum_{j=1}^{m} \alpha_{lj} s_j \omega_j)} \Big)^{\delta_i}. \tag{4.14}$$

The corresponding partial pseudo log likelihood function is given by

$$l_{ps}(\beta) = \sum_{i=1}^{n} \delta_i \Big(\beta \sum_{j=1}^{m} \alpha_{ij} s_j \omega_j - \ln \big(\sum_{l \in R(t_i)} \exp(\beta \sum_{j=1}^{m} \alpha_{lj} s_j \omega_j) \big) \Big).$$

This function can be maximized simultaneously with respect to β and $\boldsymbol{\omega}$ using AMPL in combination with the NEOS solvers. Simulations will have to show, whether this approach is justified and, therefore, an admissible alternative to the computationally more intensive maximization of the log likelihood (4.11).

4.4.5 Further comments

Whereas the likelihood function is derived given the observations $(Y_i, Z_{l_i}, Z_{l_i}, \delta_i)$, $i = 1, \ldots, n$, there is no such equivalent for the proposed pseudo likelihood function (4.14). More theoretical reinforcement is needed to find out, whether there is any theoretical derivation of this function and not only the heuristic approach.

Only right-censored survival times have been considered, nonetheless, left- or interval-censoring can also be considered. However, the resulting log likelihood functions to be maximized become more cumbersome. If further covariates are included in the model, summarized in the vector $\boldsymbol{X}$, in all the likelihood functions above, the term $\exp(\beta \boldsymbol{z})$ has to be expanded to $\exp(\beta \boldsymbol{z} + \boldsymbol{\kappa}' \boldsymbol{x})$.