Parametric proportional hazards and accelerated failure time models

Göran Broström

February 16, 2009

Abstract

A unified implementation of parametric proportional hazards (PH) and accelerated failure time (AFT) models for right-censored or interval-censored and left-truncated data is described. The description here is valid for time-constant covariates, but the necessary modifications for handling time-varying covariates are implemented in eha. Note that only piecewise constant time variation is handled.

1 Introduction

There is a need for software for analyzing parametric proportional hazards (PH) and accelerated failure time (AFT) data, that are right or interval censored and left truncated.

2 The proportional hazards model

We define proportional hazards models in terms of an expansion of a given survivor function $S_0$,

$$s_\theta(t; \mathbf{z}) = \{S_0(g(t, \theta))\}^{\exp(\mathbf{z} \beta)},$$  

(1)

where $\theta$ is a parameter vector used in modeling the baseline distribution, $\beta$ is the vector of regression parameters, and $g$ is a positive function, which helps defining a parametric family of baseline survivor functions through

$$S(t; \theta) = S_0(g(t, \theta)), \quad t > 0, \quad \theta \in \Theta.$$  

(2)
With \( f_0 \) and \( h_0 \) defined as the density and hazard functions corresponding to \( S_0 \), respectively, the density function corresponding to \( S \) is
\[
f(t; \theta) = -\frac{\partial}{\partial t} S(t, \theta) \\
= -\frac{\partial}{\partial t} S_0(g(t, \theta)) \\
= g_t(t, \theta) f_0(g(t, \theta)),
\]
where
\[
g_t(t, \theta) = \frac{\partial}{\partial t} g(t, \theta).
\]
Correspondingly, the hazard function is
\[
h(t; \theta) = \frac{f(t; \theta)}{S(t; \theta)} \\
= g_t(t, \theta) h_0(g(t, \theta)).
\]
So, the proportional hazards model is
\[
\lambda_\theta(t; z) = h(t; \theta) \exp(z_\beta) \\
= g_t(t, \theta) h_0(g(t, \theta)) \exp(z_\beta),
\]
corresponding to (1).

2.1 Data and the likelihood function

Given left truncated and right or interval censored data \((s_i, t_i, u_i, d_i, z_i), i = 1, \ldots, n\) and the model (4), the likelihood function becomes
\[
L((\theta, \beta); (s, t, u, d), Z) = \prod_{i=1}^{n} \left\{ \left( h(t_i; \theta) \exp(z_i \beta) \right)^{I(d_i=1)} \right. \\
\times \left( S(t_i; \theta)^{\exp(z_i \beta)} \right)^{I(d_i \neq 2)} \\
\times \left. \left( S(t_i; \theta)^{\exp(z_i \beta)} - S(u_i; \theta)^{\exp(z_i \beta)} \right)^{I(d_i=2)} \right\} \\
\times S(s_i; \theta)^{-\exp(z_i \beta)}
\]
Here, for \( i = 1, \ldots, n \), \( s_i < t_i \leq u_i \) are the left truncation and exit intervals, respectively, \( d_i = 0 \) means that \( t_i = u_i \) and right censoring at \( u_i \), \( d_i = 1 \) means that \( t_i = u_i \) and an event at \( u_i \), and \( d_i = 2 \) means that \( t_i < u_i \) and an event occurs in the interval \((t_i, u_i)\) (interval censoring) and \( z_i = (z_{i1}, \ldots, z_{ip}) \) is a vector of explanatory variables with corresponding parameter vector \( \beta = (\beta_1, \ldots, \beta_p), i = 1, \ldots, n \).
From (5) we now get the log likelihood and the score vector in a straightforward manner.

\[
\ell((\theta, \beta); (s, t, u, d), Z) = \sum_{i: d_i = 1} \left\{ \log h(t_i; \theta) + z_i \beta \right\} \\
+ \sum_{i: d_i \neq 2} e^{z_i \beta} \log S(u_i; \theta) \\
+ \sum_{i: d_i = 2} \log \left\{ S(t_i; \theta)^{z_i \beta} - S(u_i; \theta)^{z_i \beta} \right\} \\
- \sum_{i=1}^n e^{z_i \beta} \log S(s_i; \theta)
\]

and (in the following we drop the long argument list to \(\ell\)), for the regression parameters \(\beta\),

\[
\frac{\partial}{\partial \beta_j} \ell = \sum_{i: d_i = 1} z_{ij} \\
+ \sum_{i: d_i \neq 2} z_{ij} e^{z_i \beta} \log S(t_i; \theta) \\
+ \sum_{i: d_i = 2} z_{ij} e^{z_i \beta} \frac{S(t_i; \theta)^{z_i \beta} \log S(t_i; \theta) - S(u_i; \theta)^{z_i \beta} \log S(u_i; \theta)}{S(t_i; \theta)^{z_i \beta} - S(u_i; \theta)^{z_i \beta}} \\
- \sum_{i=1}^n z_{ij} e^{z_i \beta} \log S(s_i; \theta), \quad j = 1, \ldots, p,
\]

and for the “baseline” parameters \(\theta\), in vector form,

\[
\frac{\partial}{\partial \theta} \ell = \sum_{i: d_i = 1} h_\theta(t_i, \theta) \\
+ \sum_{i: d_i \neq 2} e^{z_i \beta} \frac{S_\theta(t_i; \theta)}{S(t_i; \theta)} \\
+ \sum_{i: d_i = 2} e^{z_i \beta} \frac{S(t_i; \theta)^{z_i \beta - 1} S_\theta(t_i; \theta) - S(u_i; \theta)^{z_i \beta - 1} S_\theta(u_i; \theta)}{S(t_i; \theta)^{z_i \beta} - S(u_i; \theta)^{z_i \beta}} \\
- \sum_{i=1}^n e^{z_i \beta} \frac{S_\theta(s_i; \theta)}{S(s_i; \theta)}.
\]

From (3),

\[
h_\theta(t, \theta) = \frac{\partial}{\partial \theta} h(t, \theta) \\
= g_\theta(t, \theta) h_0(g(t, \theta)) + g_0(t, \theta) g_\theta(t, \theta) h'_0(g(t, \theta)),
\]
and, from (2),
\[
S_\theta(t; \theta) = \frac{\partial}{\partial \theta} S(t; \theta) = \frac{\partial}{\partial \theta} S_0(g(t, \theta)) = -g_\theta(t, \theta)f_0(g(t, \theta)).
\]  

(10)

For estimating standard errors, the observed information (the negative of the hessian) is useful. However, instead of the error-prone and tedious work of calculating analytic second-order derivatives, we will rely on approximations by numerical differentiation.

3 The shape–scale families

From (1) we get a shape–scale family of distributions by choosing \( \theta = (p, \lambda) \) and
\[
g(t, (p, \lambda)) = \left(\frac{t}{\lambda}\right)^p, \quad t \geq 0; \quad p, \lambda > 0.
\]
However, for reasons of efficient numerical optimization and normality of parameter estimates, we use the parametrisation \( p = \exp(\gamma) \) and \( \lambda = \exp(\alpha) \), thus redefining \( g \) to
\[
g(t, (\gamma, \alpha)) = \left(\frac{t}{\exp(\alpha)}\right)^{\exp(\gamma)}, \quad t \geq 0; \quad -\infty < \gamma, \alpha < \infty.
\]  

(11)

For the calculation of the score and hessian of the log likelihood function, we need some partial derivatives of \( g \). They are found in an appendix.

3.1 The Weibull family of distributions

The Weibull family of distributions is obtained by
\[
S_0(t) = \exp(-t), \quad t \geq 0,
\]
leading to
\[
f_0(t) = \exp(-t), \quad t \geq 0,
\]
and
\[
h_0(t) = 1, \quad t \geq 0.
\]
We need some first and second order derivatives of \( f \) and \( h \), and they are particularly simple in this case, for \( h \) they are both zero, and for \( f \) we get
\[
f_0'(t) = -\exp(-t), \quad t \geq 0.
\]
3.2 The EV family of distributions

The EV (Extreme Value) family of distributions is obtained by setting

\[ h_0(t) = \exp(t), \quad t \geq 0, \]

leading to

\[ S_0(t) = \exp\{-\exp(t) - 1\}, \quad t \geq 0, \]

The rest of the necessary functions are easily derived from this.

3.3 The Gompertz family of distributions

The Gompertz family of distributions is given by

\[ h(t) = \tau \exp(t/\lambda), \quad t \geq 0; \quad \tau, \lambda > 0. \]

This family is not directly possible to generate from the described shape-scale models, so it is treated separately by direct maximum likelihood.

3.4 Other families of distributions

Included in the \textit{eha} package are the lognormal and the loglogistic distributions as well.

4 The accelerated failure time model

In the description of this family of models, we generate a shape-scale family of distributions as defined by the equations (2) and (11). We get

\[ S(t; (\gamma, \alpha)) = S_0\left(g(t, (\gamma, \alpha))\right) \]

\[ = S_0\left(\left\{\frac{t}{\exp(\alpha)}\right\}^{\exp(\gamma)}\right), \quad t > 0, \quad -\infty < \gamma, \alpha < \infty. \]  

(12)

Given a vector \( z = (z_1, \ldots, z_p) \) of explanatory variables and a vector of corresponding regression coefficients \( \beta = (\beta_1, \ldots, \beta_p) \), the AFT regression model is defined by

\[ S(t; (\gamma, \alpha, \beta)) = S_0\left(g(t \exp(z\beta), (\gamma, \alpha))\right) \]

\[ = S_0\left(\left\{\frac{t \exp(z\beta)}{\exp(\alpha)}\right\}^{\exp(\gamma)}\right) \]

\[ = S_0\left(\left\{\frac{t}{\exp(\alpha - z\beta)}\right\}^{\exp(\gamma)}\right) \]

\[ = S_0\left(g(t, (\gamma, \alpha - z\beta))\right), \quad t > 0. \]

(13)
So, by defining $\theta = (\gamma, \alpha - z\beta)$, we are back in the framework of Section 2. We get

$$f(t; \theta) = g(t, \theta)f_0(g(t, \theta))$$

and

$$h(t; \theta) = g(t, \theta)h_0(g(t, \theta)),$$

defining the AFT model generated by the survivor function $S_0$ and corresponding density $f_0$ and hazard $h_0$.

4.1 Data and the likelihood function

Given left truncated and right or interval censored data $(s_i, t_i, u_i, d_i, z_i), i = 1, \ldots, n$ and the model (14), the likelihood function becomes

$$L((\gamma, \alpha, \beta); (s, t, u, d), Z) = \prod_{i=1}^{n} \left\{ \begin{array}{l}
\log h(t_i; \theta_i) I(d_i=1) \\
\log S(t_i; \theta_i) I(i\neq 2) \\
\log \left( S(t_i; \theta_i) - S(u_i; \theta_i) \right) I(d_i=2) \\
\log S(s_i; \theta_i)^{-1} \end{array} \right\}$$

(15)

Here, for $i = 1, \ldots, n$, $s_i < t_i \leq u_i$ are the left truncation and exit intervals, respectively, $d_i = 0$ means that $t_i = u_i$ and right censoring at $t_i$, $d_i = 1$ means that $t_i = u_i$ and an event at $t_i$, and $d_i = 2$ means that $t_i < u_i$ and an event occurs in the interval $(t_i, u_i)$ (interval censoring), and $z_i = (z_{i1}, \ldots, z_{ip})$ is a vector of explanatory variables with corresponding parameter vector $\beta = (\beta_1, \ldots, \beta_p)$, $i = 1, \ldots, n$.

From (15) we now get the log likelihood and the score vector in a straightforward manner.

$$\ell((\gamma, \alpha, \beta); (s, t, u, d), Z) = \sum_{i:d_i=1} \log h(t_i; \theta_i)$$

$$+ \sum_{i:d_i\neq 2} \log S(t_i; \theta_i)$$

$$+ \sum_{i:d_i=2} \log \left( S(t_i; \theta_i) - S(u_i; \theta_i) \right)$$

$$- \sum_{i=1}^{n} \log S(s_i; \theta_i)$$

and (in the following we drop the long argument list to $\ell$), for the regression parameters $\beta$, 

---

6
\[
\frac{\partial}{\partial \beta_j} \ell = \sum_{d_i=1} h_j(t_i, \theta_i) + \sum_{d_i \neq 2} S_j(t_i; \theta_i) \\
+ \sum_{d_i=2} S_j(t_i; \theta_i) - S_j(u_i; \theta_i) - \sum_{i=1}^n S_j(s_i; \theta_i) \\
= - \sum_{d_i=1} z_{ij} h_\alpha(t_i, \theta_i) - \sum_{d_i \neq 2} z_{ij} S_\alpha(t_i; \theta_i) \\
- \sum_{d_i=2} z_{ij} S_\alpha(t_i; \theta_i) - S_\alpha(u_i; \theta_i) + \sum_{i=1}^n z_{ij} S_\alpha(s_i; \theta_i)
\]

and for the “baseline” parameters \(\gamma\) and \(\alpha\),

\[
\frac{\partial}{\partial \gamma} \ell = \sum_{i:d_i=1} h_\gamma(t_i, \theta_i) + \sum_{i:d_i \neq 2} S_\gamma(t_i; \theta_i) \\
+ \sum_{i:d_i=2} S_\gamma(t_i; \theta_i) - S_\gamma(u_i; \theta_i) - \sum_{i=1}^n S_\gamma(s_i; \theta_i),
\]

and

\[
\frac{\partial}{\partial \alpha} \ell = \sum_{i:d_i=1} h_\alpha(t_i, \theta_i) + \sum_{i:d_i \neq 2} S_\alpha(t_i; \theta_i) \\
+ \sum_{i:d_i=2} S_\alpha(t_i; \theta_i) - S_\alpha(u_i; \theta_i) - \sum_{i=1}^n S_\alpha(s_i; \theta_i).
\]

Here, from (3),

\[
h_\gamma(t, \theta_i) = \frac{\partial}{\partial \gamma} h(t, \theta_i) \\
= g_\gamma(t, \theta_i) h_0(g(t, \theta_i)) + g_\gamma(t, \theta_i) h_0'(g(t, \theta_i)),
\]

\[
h_\alpha(t, \theta_i) = \frac{\partial}{\partial \alpha} h(t, \theta_i) \\
= g_\alpha(t, \theta_i) h_0(g(t, \theta_i)) + g_\alpha(t, \theta_i) h_0'(g(t, \theta_i)),
\]

and

\[
h_j(t, \theta_i) = \frac{\partial}{\partial \beta_j} h(t, \theta_i) = \frac{\partial}{\partial \alpha} h(t, \theta_i) \frac{\partial}{\partial \beta_j} (\alpha - z_i \beta) \\
= -z_{ij} h_\alpha(t, \theta_i), \quad j = 1, \ldots, p.
\]
Similarly, from (2) we get

\[ S_\gamma(t; \theta_i) = \frac{\partial}{\partial \gamma} S(t; \theta_i) = \frac{\partial}{\partial \gamma} S_0(g(t, \theta_i)) = -g_\gamma(t, \theta_i) f_0(g(t, \theta_i)), \]

\[ S_\alpha(t; \theta_i) = \frac{\partial}{\partial \alpha} S(t; \theta_i) = \frac{\partial}{\partial \alpha} S_0(g(t, \theta_i)) = -g_\alpha(t, \theta_i) f_0(g(t, \theta_i)). \]

and

\[ S_j(t; \theta_i) = \frac{\partial}{\partial \beta_j} S(t; \theta_i) = \frac{\partial}{\partial \alpha} S_0(g(t, \theta_i)) \frac{\partial}{\partial \beta_j} (\alpha - z_i \beta) = -z_{ij} S_\alpha(t; \theta_i), \quad j = 1, \ldots, p. \]

For estimating standard errors, the observed information (the negative of the hessian) is useful, so

\[ -\frac{\partial^2}{\partial \beta_j \partial \beta_m} \ell = \sum_{i: d_i = 1} z_{ij} z_{im} \left\{ \frac{h_{aa}(t_i, \theta_i)}{h(t_i, \theta_i)} - \left( \frac{h_\alpha(t_i, \theta_i)}{h(t_i, \theta_i)} \right)^2 \right\} \]

\[ -\sum_{i: i \neq 2} z_{ij} z_{im} \left\{ \frac{S_{aa}(t_i, \theta_i)}{S(t_i, \theta_i)} - \left( \frac{S_\alpha(t_i, \theta_i)}{S(t_i, \theta_i)} \right)^2 \right\} \]

\[ -\sum_{i = 2} \sum_{i: \neq 2} z_{ij} z_{im} \left\{ \frac{S_{aa}(t_i, \theta_i) - S_{aa}(u_i, \theta_i)}{S(t_i, \theta_i) - S(u_i, \theta_i)} - \left( \frac{S_\alpha(t_i, \theta_i) - S_\alpha(u_i, \theta_i)}{S(t_i, \theta_i) - S(u_i, \theta_i)} \right)^2 \right\} \]

\[ +\sum_{i = 1} \sum_{i: \neq 2} z_{ij} z_{im} \left\{ \frac{S_{aa}(s_i, \theta_i)}{S(s_i, \theta_i)} - \left( \frac{S_\alpha(s_i, \theta_i)}{S(s_i, \theta_i)} \right)^2 \right\}, \quad j, m = 1, \ldots, p. \]
and

\[- \frac{\partial^2}{\partial \beta_j \partial \tau} \ell = \sum_{i: i \neq 1} z_{ij} \left\{ \frac{h_{\alpha \tau}(t_i, \theta_i)}{h(t_i, \theta_i)} - \frac{h_{\alpha}(t_i, \theta_i)}{h^2(t_i, \theta_i)} \right\} \]

\[+ \sum_{i : i \neq 2} z_{ij} \left\{ \frac{S_{\alpha \tau}(t_i, \theta_i) S_{\tau}(t_i, \theta_i)}{S(t_i, \theta_i)} - \frac{S_{\alpha}(t_i, \theta_i)}{S^2(t_i, \theta_i)} \right\} \]

\[- \frac{\partial^2}{\partial \tau \partial \tau'} \ell = - \sum_{i : i \neq 1} z_{ij} \left\{ \frac{h_{\tau \tau'}(t_i, \theta_i)}{h(t_i, \theta_i)} - \frac{h_{\tau}(t_i, \theta_i)}{h^2(t_i, \theta_i)} \right\} \]

\[+ \sum_{i : i \neq 2} z_{ij} \left\{ \frac{S_{\tau \tau'}(t_i, \theta_i) S_{\tau}(t_i, \theta_i)}{S(t_i, \theta_i)} - \frac{S_{\tau}(t_i, \theta_i)}{S^2(t_i, \theta_i)} \right\} \]

\[- \frac{\partial^2}{\partial \tau \partial \tau'} \ell = - \sum_{i : i \neq 2} \left\{ \frac{S_{\tau \tau'}(t_i, \theta_i) S_{\tau'}(t_i, \theta_i)}{S(t_i, \theta_i)} - \frac{S_{\tau'}(t_i, \theta_i)}{S^2(t_i, \theta_i)} \right\} \]

\[+ \sum_{i = 1}^n \left\{ \frac{S_{\tau \tau'}(s_i, \theta_i)}{S(s_i, \theta_i)} - \frac{S_{\tau'}(s_i, \theta_i)}{S^2(s_i, \theta_i)} \right\} \]

\[= 1, \ldots, p; \quad \tau = \gamma, \alpha, \tau' = \gamma, \alpha, (\gamma, \alpha), (\alpha, \alpha).\]
The second order partial derivatives $h_{\tau\tau'}$ and $S_{\tau\tau'}$ are

$$h_{\tau\tau'}(t, \theta) = \frac{\partial}{\partial \tau'} h_{\tau}(t, \theta)$$

$$= g_{\tau\tau'}(t, \theta) h_0(g(t, \theta)) + g_{\tau\tau}(t, \theta) g_{\tau'}(t, \theta) h_0'(g(t, \theta))$$

$$+ g_{\tau}(t, \theta) g_{\theta\theta}(t, \theta) h_0'(g(t, \theta))$$

$$+ g_{\tau\tau}(t, \theta) g_{\theta}(t, \theta) h_0'(g(t, \theta))$$

$$= h_0(g(t, \theta)) g_{\tau\tau'}(t, \theta)$$

$$+ h_0'(g(t, \theta)) \{ g_{\tau}(t, \theta) g_{\theta\theta}(t, \theta)$$

$$+ g_{\tau\tau}(t, \theta) g_{\theta}(t, \theta)$$

$$+ g_{\tau}(t, \theta) g_{\theta}(t, \theta) \}$$

$$+ h_0''(g(t, \theta)) g_{\tau}(t, \theta) g_{\theta}(t, \theta) g_{\tau'}(t, \theta)$$

$$(\tau, \tau') = (\gamma, \gamma), (\gamma, \lambda), (\lambda, \lambda), (16)$$

and from (10),

$$S_{\tau\tau'}(t, \theta) = \frac{\partial}{\partial \tau'} S_{\tau}(t, \theta)$$

$$= - \{ g_{\tau\tau'}(t, \theta) f_0(g(t, \theta)) + g_{\tau}(t, \theta) g_{\tau'}(t, \theta) f_0'(g(t, \theta)) \}$$

$$(\tau, \tau') = (\gamma, \gamma), (\gamma, \lambda), (\lambda, \lambda). (17)$$

### A Some partial derivatives

The function (see (11))

$$g(t, (\gamma, \alpha)) = \left( \frac{t}{\exp(\alpha)} \right)^{\exp(\gamma)} \quad t \geq 0; \quad -\infty < \gamma, \alpha < \infty. (18)$$

has the following partial derivatives:

$$g_t(t, (\gamma, \alpha)) = \frac{\exp(\gamma)}{t} g(t, (\gamma, \alpha)), \quad t > 0$$

$$g_{\gamma}(t, (\gamma, \alpha)) = g(t, (\gamma, \alpha)) \log \{ g(t, (\gamma, \alpha)) \}$$

$$g_{\alpha}(t, (\gamma, \alpha)) = - \exp(\gamma) g(t, (\gamma, \alpha))$$
\[ g_{t\gamma}(t, (\gamma, \alpha)) = g_t(t, (\gamma, \alpha)) + \frac{\exp(\gamma)}{t} g_{\gamma}(t, (\gamma, \alpha)), \quad t > 0 \]
\[ g_{t\alpha}(t, (\gamma, \alpha)) = -\exp(\gamma) g_t(t, (\gamma, \alpha)), \quad t > 0 \]
\[ g_{\gamma^2}(t, (\gamma, \alpha)) = g_\gamma(t, (\gamma, \alpha)) \{ 1 + \log g(t, (\gamma, \alpha)) \} \]
\[ g_{\gamma\alpha}(t, (\gamma, \alpha)) = g_\alpha(t, (\gamma, \alpha)) \{ 1 + \log g(t, (\gamma, \alpha)) \} \]
\[ g_{\alpha^2}(t, (\gamma, \alpha)) = -\exp(\gamma) g_\alpha(t, (\gamma, \alpha)) \]

\[ g_{t\gamma^2}(t, (\gamma, \alpha)) = g_{t\gamma}(t, (\gamma, \alpha)) \]
\[ + \frac{\exp(\gamma)}{t} g_\gamma(t, (\gamma, \alpha)) \{ 2 + \log g(t, (\gamma, \alpha)) \} \]
\[ g_{t\gamma\alpha}(t, (\gamma, \alpha)) = -\exp(\gamma) \{ g_t(t, (\gamma, \alpha)) + g_{t\gamma}(t, (\gamma, \alpha)) \} \]
\[ g_{t\alpha^2}(t, (\gamma, \alpha)) = -\exp(\gamma) g_{t\alpha}(t, (\gamma, \alpha)) \]

The formulas will be easier to read if we remove all function arguments, i.e., \((t, (\gamma, \alpha))\):

\[ g_t = \frac{\exp(\gamma)}{t} g, \quad t > 0 \]
\[ g_\gamma = g \log g \]
\[ g_\alpha = -\exp(\gamma) g \]
\[ g_{t\gamma} = g_t + \frac{\exp(\gamma)}{t} g_\gamma, \quad t > 0 \]
\[ g_{t\alpha} = -\exp(\gamma) g_t, \quad t > 0 \]
\[ g_{\gamma^2} = g_\gamma \{ 1 + \log g \} \]
\[ g_{\gamma\alpha} = g_\alpha \{ 1 + \log g \} \]
\[ g_{\alpha^2} = -\exp(\gamma) g_\alpha \]
\[ g_{t\gamma^2} = g_\gamma + \frac{\exp(\gamma)}{t} g_\gamma \{ 2 + \log g \}, \quad t > 0 \]
\[ g_{t\gamma\alpha} = -\exp(\gamma) \{ g_t + g_{t\gamma} \}, \quad t > 0 \]
\[ g_{t\alpha^2} = -\exp(\gamma) g_{t\alpha}, \quad t > 0 \]