Technical Appendix for `coxph.risk`
Stephanie A. Kovalchik
August 29, 2012

1 Definition

Letting m $(1, \ldots, M)$ index the number of failure types, the absolute risk of experiencing the mth event within the time interval $[t_0, t_1)$ in the presence of $M-1$ competing events is

$$\pi_m(t_0, t_1; \vec{x}) = \left[\prod_{i=1}^{M} S_i(t_0; \vec{x}_i)\right]^{-1} \int_{t_0}^{t_1} \lambda_m(u; \vec{x}_m) \prod_{i=1}^{M} S_i(u; \vec{x}_i) du. \tag{1}$$

where $\vec{x} = (\vec{x}_1, \ldots, \vec{x}_M)$ is a set of cause-specific covariate vectors, $S_m(u; x_m) = \exp(-\int_0^u \lambda_m(v; \vec{x}_m) dv)$ and $\lambda_m(u; \vec{x}_m)$ are the cause-specific survival and hazard functions given covariates $\vec{x}_m$. We assume that covariates in (1) remain fixed at their values at the beginning of the projection interval, t_0. For simplicity, the subscript in π_m, which emphasizes that the absolute risk pertains to a particular cause, will be omitted from here on.

The formulation of absolute risk given in Equation (1) can accomodate many possible hazard models. In the `coxph.risk` implementation, the hazard model for each cause follows Cox's proportional hazards model,

$$\lambda_m(t; \vec{x}_m) = \lambda_{0m}(t) \exp(\boldsymbol{\beta}'_m \vec{x}_m) \tag{2}$$

where $\lambda_{0m}(t)$ denotes the baseline hazard function at time t.

2 Estimation

Consider a cohort of $i = 1, \ldots, n$ individuals. Let $\delta_i^m(t)$ be an indicator function for the ith individual and mth event at time t, and let $y_i^m(t)$ indicate the at-risk status at time t mth event at time t, taking the value one when the ith individual experiences an event or is censored at t or later. The estimating equations for $\boldsymbol{\beta}'_m$ $(1, \ldots, M)$ are

$$\vec{U}(\boldsymbol{\beta}_m) = \sum_{i=1}^{n} \delta_i^m(t_i)\{\vec{x}_i^m - \vec{\bar{H}}(\boldsymbol{\beta}_m, t_i)\} \tag{3}$$

where $\vec{\bar{H}}(\boldsymbol{\beta}_m, t)$ is an 'average' of the risk profiles $\vec{x}^m$ among the individuals still at-risk at time t,

$$\vec{\bar{H}}(\boldsymbol{\beta}_m, t) = \frac{\sum_{i=1}^{n} y_i^m(t) \exp(\boldsymbol{\beta}'_m \vec{x}_i^m) \vec{x}_i^m}{\sum_{i=1}^{n} y_i^m(t) \exp(\boldsymbol{\beta}'_m \vec{x}_i^m)}. \tag{4}$$

Standard optimization algorithms can be used to obtain the solution $\hat{\boldsymbol{\beta}}_m$ to the estimating equations in (3).

When no distributional assumption is made for λ_{0m}, the estimator for the cause-specific risk of the primary event within the interval $[t_0, t_1)$, given $\vec{x}$, is

$$\hat{\pi}(t_0, t_1; \vec{x}) = \left[\prod_{i=1}^{M} \hat{S}_{0i}(t_0)^{\exp(\hat{\boldsymbol{\beta}}'_i \vec{x}^i)}\right]^{-1} \exp(\hat{\boldsymbol{\beta}}'_1 \vec{x}^1) \sum_{t_0 \le u < t_1} \hat{\lambda}_{01}(u) \prod_{i=1}^{M} \hat{S}_{0i}(u)^{\exp(\hat{\boldsymbol{\beta}}'_i \vec{x}^i)}, \tag{5}$$

where $\hat{S}_{0i}(u)$ is the cause-specific baseline survival function and $\hat{\lambda}_{01}(u)$ the primary-event baseline hazard function at time u. A semiparametric weighted Nelson-Aalen estimator (Aalen 1978) for the cause-specific baseline hazard function is

$$\hat{\lambda}_{0m}(t) = \frac{\sum_{i=1}^{n} y_i^m(t)\delta_i^m(t)}{\sum_{i=1}^{n} y_i^m(t)\exp(\hat{\boldsymbol{\beta}}_m^{\prime}\vec{x}_i^m)}, \tag{6}$$

which uses Breslow's method for handling ties (Breslow 1974). The cause-specific baseline survival at time t is estimated as

$$\hat{S}_{0m}(t) = \exp(-\sum_{u^m \leq t} \hat{\lambda}_{0m}(u_i^m)), \tag{7}$$

with u_i^m denoting the observed event times for the mth event type.

For both the piecewise and semiparametric approaches, given a baseline survival estimate, the survival to time t for an individual with risk profile $\vec{x}^m$ is

$$\hat{S}_m(t;\vec{x}^m) = \hat{S}_{0m}(t)^{\exp(\hat{\boldsymbol{\beta}}_m^{\prime}\vec{x}^m)}. \tag{8}$$

3 Variance

Denote the N^m ordered observed event times occuring within $[t_0, t_1)$ for the mth cause as $u_1^m < u_2^m < \cdots < u_{N^m}^m$. In terms of these event times, Equation (1) becomes

$$\hat{\pi}(t_0, t_1; \vec{x}) = \exp(\hat{\boldsymbol{\beta}}_1^{\prime}\vec{x}^1) \sum_{i=1}^{N^1} \hat{\lambda}_{01}(u_i^1) \prod_{j=1}^{M} \left(\hat{S}_{0j}(u_i^1)/\hat{S}_{0j}(u_1^1)\right)^{\exp(\hat{\boldsymbol{\beta}}_j^{\prime}\vec{x}^j)} = \sum_{i=1}^{N^1} \hat{\pi}(u_i^1). \tag{9}$$

with $\hat{\pi}(u_i^1) = \exp(\hat{\boldsymbol{\beta}}_1^{\prime}\vec{x}^1)\hat{\lambda}_{01}(u_i^1)\prod_{j=1}^{M}\left(\hat{S}_{0j}(u_i^1)/\hat{S}_{0j}(u_1^1)\right)^{\exp(\hat{\boldsymbol{\beta}}_j^{\prime}\vec{x}^j)}$.

We determine the derivative and deviates for each component of (9). For the $\hat{\boldsymbol{\beta}}_j$, the derivate is

$$\frac{\partial\hat{\pi}(t_0, t_1; \vec{x})}{\partial\hat{\boldsymbol{\beta}}_j} = \vec{x}^j \left[\hat{\pi}(t_0, t_1; \vec{x}) + \exp(\hat{\boldsymbol{\beta}}_j^{\prime}\vec{x}^j) \sum_{i=1}^{N_1} \log\left(\hat{S}_{0j}(u_i^1)/\hat{S}_{0j}(u_1^1)\right) \hat{\pi}(u_i^1)\right],$$

when $j = 1$ and

$$\frac{\partial\hat{\pi}(t_0, t_1; \vec{x})}{\partial\hat{\boldsymbol{\beta}}_j} = \vec{x}^j \exp(\hat{\boldsymbol{\beta}}_j^{\prime}\vec{x}^j) \sum_{i=1}^{N_1} \log\left(\hat{S}_{0j}(u_i^1)/\hat{S}_{0j}(u_1^1)\right) \hat{\pi}(u_i^1)$$

for competing causes. The Taylor deviates for each $\hat{\boldsymbol{\beta}}_m$ are

$$\Delta_i\{\hat{\boldsymbol{\beta}}_m\} = \mathcal{H}(\hat{\boldsymbol{\beta}}_m)^{-1} \sum_{j=1}^{n} \delta_j^m(t_i)\{\vec{x}_j^m - \bar{\bar{\vec{H}}}(\hat{\boldsymbol{\beta}}_m, t_j)\}. \tag{10}$$

The derivatives for the baseline hazard components are

$$\frac{\partial\hat{\pi}(t_0, t_1; \vec{x})}{\partial\hat{\lambda}_{01}(u_i^1)} = \hat{\lambda}_{01}(u_i^1)^{-1}\hat{\pi}(u_i^1). \tag{11}$$

The Taylor deviates for the baseline hazard of cause m at observed event time t are

$$\Delta_i\{\hat{\lambda}_{0m}(t)\} = \frac{\partial \hat{\lambda}_{0m}(t)}{\partial N_m(t)} \Delta_i\{N_m(t)\} + \frac{\partial \hat{\lambda}_{0m}(t)}{\partial G_m(t)} \Delta_i\{G_m(t)\}, \tag{12}$$

where

$$N_m(t) = \sum_{i=1}^{n} y_i^m(t) \delta_i^m(t)$$

and

$$G_m(t) = \sum_{i=1}^{n} y_i^m(t) \exp(\hat{\boldsymbol{\beta}}_m' \vec{x}_i^m).$$

In terms of these quantities, the Taylor deviates are

$$\Delta_i\{\hat{\lambda}_{0m}(t)\} = G_m(t)^{-1}(y_i^m(t)\delta_i^m(t) - \hat{\lambda}_{0m}(t)\Delta_i\{G_m(t)\}) \tag{13}$$

with

$$\begin{aligned} \Delta_i\{G_m(t)\} = \quad & y_i^m(t) \exp(\hat{\boldsymbol{\beta}}_m' \vec{x}_i^m) \\ & + \left[\textstyle\sum_{j=1}^{n} \vec{x}_j y_j^m(t) \exp(\hat{\boldsymbol{\beta}}_m' \vec{x}_j^m) \right] \Delta_i\{\hat{\boldsymbol{\beta}}_m\}. \end{aligned}$$

The final components are the survival functions. The derivatives for each $\hat{S}_{0m}(u_j^1)$ are

$$\frac{\partial \hat{\pi}(t_0, t_1; \vec{x})}{\partial \hat{S}_{0m}(u_j^1)} = sgn(m) \exp(\hat{\boldsymbol{\beta}}_m' \vec{x}^m) \hat{S}_{0m}(u_j^1)^{-1} \hat{\pi}(u_j^1) \tag{14}$$

where $sgn(1) = -1$ and is one otherwise. Given the semiparametric estimate,

$$\hat{S}_{0m}(t) = \exp(-\sum_{u_i^m \le t} \hat{\lambda}_{0m}(u_i^m)), \tag{15}$$

the Taylor deviates for the baseline survival up to time u_j^1 for the mth risk type are

$$\Delta_i\{\hat{S}_{0m}(u_j^1)\} = -\hat{S}_{0m}(u_j^1) \sum_{u_n^m \le u_j^1} \Delta_i\{\hat{\lambda}_{0m}(u_n^m)\}. \tag{16}$$

Combining these results, the expression for the Taylor deviates of $\hat{\pi}(t_0, t_1; \vec{x})$ are

$$\begin{aligned} \Delta_i\{\hat{\pi}(t_0, t_1; \vec{x})\} = \quad & \textstyle\sum_{m=1}^{M} \frac{\hat{\pi}(t_0, t_1; \vec{x})}{\partial \hat{\boldsymbol{\beta}}_m} \Delta_i\{\hat{\boldsymbol{\beta}}_m\} + \sum_{j=1}^{N_1} \frac{\hat{\pi}(t_0, t_1; \vec{x})}{\partial \hat{\lambda}_{01}(u_j^1)} \Delta_i\{\hat{\lambda}_{01}(u_l^1)\} \\ & + \textstyle\sum_{j=1}^{N_1} \sum_{m=1}^{M} \frac{\hat{\pi}(t_0, t_1; \vec{x})}{\partial \hat{S}_{0m}(u_j^1)} \Delta_i\{\hat{S}_{0m}(u_j^1)\}. \end{aligned}$$