Algorithm for Cox Models in Clustered Survival Analysis

1 Model and Data Structure

Let T_{ij} denote survival time and C_{ij} be censoring time for subject j in cluster i. Let $Z_{ij} = (Z_{1ij}, ..., Z_{dij})^T$ be d-dimensional covariates. The Cox proportional hazards models assume that cumulative hazards for T_{ij} given Z_{ij} and cluster-specific random effect b_i is given by

$$\Lambda_{ij}(t|Z_{ij}, b_i) = \Lambda(t)e^{Z_{ij}^T\beta + b_i},$$

where $\Lambda(t)$ is an increasing and unknown cumulative baseline function with $\Lambda(0) = 0$, β is the coefficient of Z_{ij}, and b_i is assumed to follow a normal distribution with mean zero and variance σ^2.

The observed data are $(Y_{ij} = T_{ij} \wedge C_{ij}, \Delta_{ij} = I(T_{ij} \leq C_{ij}), Z_{ij}), i = 1, ..., n, j = 1, ..., n_i$ Under the coarsening at random assumption, the complete log-likelihood function concerning $\Lambda(t)$ and β is given

$$\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left[\Delta_{ij}\log\Lambda'(Y_{ij}) + \Delta_{ij}Z_{ij}^T\beta + \Delta_{ij}b_i - \Lambda(Y_{ij})e^{Z_{ij}^T\beta + b_i}\right]$$
$$+\sum_{i=1}^{n}\left[-\frac{1}{2}\log 2\pi\sigma^2 - \frac{b_i^2}{2\sigma^2}\right],$$

where $\Lambda'(t)$ is the derivative of $\Lambda(t)$.

2 Algorithm for Calculating NPMLE

In the NPMLE, Λ is assumed to be a step function with jump sizes at observed event Y_{ij}. We define q_{ij} be the jump size of $\Lambda(t)$ at Y_{ij} and $Q_{ij} = \sum_{k,l} I(Y_{kl} \leq Y_{ij})q_{kl}$.

We propose the following EM algorithm for calculating the NPMLE.

Initial Step. We let $\beta = 0$, $\sigma^2 = 1$, and $q_{ij} = \Delta_{ij}/m$, where $m = \sum_{i,j}\Delta_{ij}$.

E-Step. For each $i = 1, ..., n$ and $j = 1, ..., n_i$, we then compute

$$\widehat{w}_{1ij} = \frac{\int e^{b_i} \exp\{-\frac{b_i^2}{2\sigma^2}\} f(b_i) db_i}{\int \exp\{-\frac{b_i^2}{2\sigma^2}\} f(b_i) db_i},$$

where

$$f(b_i) = \prod_{j=1}^{n_i} \left[e^{b_i} \exp\{-Q_{ij} e^{Z_{ij}^T \beta + b_i}\} \right].$$

Particularly, $\widehat{w}_{1ij}$ can be approximated using the Gaussian-Hermite quadrature: suppose $(x_1, ..., x_s)$ and $(v_1, ..., v_s)$ be the quadratures and the corresponding weights; then

$$\widehat{w}_{1ij} = \frac{\sum_{k=1}^{s} e^{\sqrt{2\sigma^2} x_k} f(\sqrt{2\sigma^2} x_k) v_k}{\sum_{k=1}^{s} f(\sqrt{2\sigma^2} x_k) v_k}.$$

M-step. We solve the following equation to updated β using one-step Newton-Raphson method:

$$S(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{n_i} \Delta_{ij} \left[Z_{ij} - \frac{\sum_{k=1}^{n} \sum_{l=1}^{n_k} I(Y_{kl} \geq Y_{ij}) Z_{kl} e^{Z_{kl}^T \beta} \widehat{w}_{1kl}}{\sum_{k=1}^{n} \sum_{l=1}^{n_k} I(Y_{kl} \geq Y_{ij}) e^{Z_{kl}^T \beta} \widehat{w}_{1kl}} \right].$$

That is, the updated β should be equal to $\beta - \Sigma^{-1} S(\beta)$, where

$$\Sigma = -\sum_{i=1}^{n} \sum_{j=1}^{n_i} \frac{\sum_{k=1}^{n} \sum_{l=1}^{n_k} I(Y_{kl} \geq Y_{ij}) Z_{kl} Z_{kl}^T e^{Z_{kl}^T \beta} \widehat{w}_{1kl}}{\sum_{k=1}^{n} \sum_{l=1}^{n_k} I(Y_{kl} \geq Y_{ij}) e^{Z_{kl}^T \beta} \widehat{w}_{1kl}} + \sum_{i=1}^{n} \left[\frac{\sum_{k=1}^{n} \sum_{l=1}^{n_k} I(Y_{kl} \geq Y_{ij}) Z_{kl} e^{Z_{kl}^T \beta} \widehat{w}_{1kl}}{\sum_{k=1}^{n} \sum_{l=1}^{n_k} I(Y_{kl} \geq Y_{ij}) e^{Z_{kl}^T \beta} \widehat{w}_{1kl}} \right]^{\otimes 2},$$

where $a^{\otimes 2} = aa^T$. Using the updated β, we calculate the updated q_{ij} as

$$q_{ij} = \frac{\Delta_{ij}}{\sum_{k=1}^{n} \sum_{l=1}^{n_k} I(Y_{kl} \geq Y_{ij}) e^{Z_{kl}^T \beta} \widehat{w}_{1kl}}.$$

We update σ^2 by

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} \widehat{\nu}_i,$$

where

$$\widehat{\mu}_i = \frac{\int b_i^2 \exp\{-\frac{b_i^2}{2\sigma^2}\} f(b_i) db_i}{\int \exp\{-\frac{b_i^2}{2\sigma^2}\} f(b_i) db_i},$$

which can also be computed using the Gaussian-Hermite quadrature.

Finally, we iterate between the E-step and M-step till convergence. We denote the final estimate as $\widehat{\beta}, \widehat{\sigma}^2, \widehat{q}_{ij}, i = 1, ..., n, j = 1, ..., n_i$. We let $y_1, ..., y_m$ be the observed Y_{ij}'s with $\Delta_{ij} = 1$ and let the estimated jump sizes associated with these times be $\widehat{p}_1, ..., \widehat{p}_m$.

3 Variance Estimation

We use the Louis formula for calculating the variance for $(\widehat{\beta}, \widehat{p}_1, ..., \widehat{p}_m, \widehat{\sigma}^2,)$.

For each subject $i = 1, .., n$, we define

$$S_{1ij} = \begin{pmatrix} \Delta_{ij} Z_{ij} \\ \Delta_{ij} I(Y_{ij} = y_1)/\widehat{p}_1 \\ \vdots \\ \Delta_{ij} I(Y_{ij} = y_m)/\widehat{p}_m \end{pmatrix}$$

and

$$S_{2ij} = \begin{pmatrix} \widehat{Q}_{ij} e^{Z_{ij}^T \widehat{\beta}} Z_{ij} \\ I(Y_{ij} \geq y_1) e^{Z_{ij}^T \widehat{\beta}} \\ \vdots \\ I(Y_{ij} \geq y_m) e^{Z_{ij}^T \widehat{\beta}} \end{pmatrix}.$$

We then calculate the score $l_{1i} = \sum_{j=1}^{n_i} \{S_{1ij} - \widehat{w}_{1ij} S_{2ij}\}$. We also obtain the matrix

$$l_{11i} = (\sum_{j=1}^{n_i} S_{1ij})^{\otimes 2} - \sum_{j=1}^{n_i} \sum_{l=1}^{n_i} \widehat{w}_{1il} S_{1ij} S_{2il}^T - \sum_{j=1}^{n_i} \sum_{l=1}^{n_i} \widehat{w}_{1il} S_{2il} S_{1ij}^T + \sum_{j=1}^{n_i} \sum_{l=1}^{n_i} \widehat{w}_{2ijl} S_{2ij} S_{2il}^T.$$

Here,

$$\widehat{w}_{2ijl} = \frac{\int e^{2b_i} \exp\{-\frac{b_i^2}{2\sigma^2}\} f(b_i) db_i}{\int \exp\{-\frac{b_i^2}{2\sigma^2}\} f(b_i) db_i}.$$

Additionally, we calculate

$$l_{2i} = \sum_{j=1}^{n_i} \begin{pmatrix} -\widehat{Q}_{ij} \widehat{w}_{1ij} e^{Z_{ij}^T \widehat{\beta}} Z_{ij} Z_{ij}^T & -I(Y_{ij} \geq y_1) \widehat{w}_{1ij} e^{Z_{ij}^T \widehat{\beta}} Z_{ij} & \dots & -(Y_{ij} \geq y_m) \widehat{w}_{1ij} e^{Z_{ij}^T \widehat{\beta}} Z_{ij} \\ -I(Y_{ij} \geq y_1) \widehat{w}_{1ij} e^{Z_{ij}^T \widehat{\beta}} Z_{ij}^T & -\Delta_{ij} I(Y_{ij} = y_1)/\widehat{p}_1^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -(Y_{ij} \geq y_m) \widehat{\xi}_{1i} e^{Z_{ij}^T \widehat{\beta}} Z_{ij}^T & 0 & \dots & -\Delta_{ij} I(Y_{ij} = y_m)/\widehat{p}_m^2 \end{pmatrix}.$$

Finally, we add the component for σ^2 to each l_{1i}, l_{11i} and l_{2i} by redefining $l_{1i} = (l_{1i}^T, -1/(2\widehat{\sigma}^2) + \widehat{\mu}_i/(2\widehat{\sigma}^4))^T$,

$$l_{11i} = \begin{pmatrix} l_{11i} & c_i \\ c_i^T & \frac{1}{4\widehat{\sigma}^4} - \frac{1}{2\widehat{\sigma}^6} \widehat{\mu}_i + \frac{1}{4\widehat{\sigma}^8} \widehat{\theta}_i \end{pmatrix},$$

where

$$c_i = (-1/(2\widehat{\sigma}^2) + \widehat{\mu}_i/(2\widehat{\sigma}^4)) \sum_{j=1}^{n_i} S_{1ij} + \frac{1}{2\widehat{\sigma}^2} \sum_{j=1}^{n_i} \widehat{w}_{1ij} S_{2ij} - \frac{1}{2\widehat{\sigma}^4} \sum_{j=1}^{n_i} \widehat{\lambda}_{ij} S_{2ij}$$

with

$$\widehat{\lambda}_{ij} = \frac{\int b_i^2 e^{b_i} \exp\{-\frac{b_i^2}{2\sigma^2}\} f(b_i) db_i}{\int \exp\{-\frac{b_i^2}{2\sigma^2}\} f(b_i) db_i}$$

and

$$\widehat{\theta}_i = \frac{\int b_i^4 \exp\{-\frac{b_i^2}{2\sigma^2}\} f(b_i) db_i}{\int \exp\{-\frac{b_i^2}{2\sigma^2}\} f(b_i) db_i},$$

and $l_{2i} = \mathrm{diag}(l_{2i}, 1/(2\widehat{\sigma}^4) - \widehat{\mu}_i/(\widehat{\sigma}^6))$.

Then the observed information matrix for $(\widehat{\beta}, \widehat{p}_1, ..., \widehat{p}_m, \widehat{\sigma}^2)$ is

$$-\sum_{i=1}^{n} l_{2i} - \sum_{i=1}^{n} \left\{ l_{11i} - l_{1i}^{\otimes 2} \right\}.$$

The covariance matrix is the inverse of this matrix.