ISyE 6404 – PH-Regression

1. Example/Data/Hand-Calculation

A simple example:

individual	X_i	δ_i	Z_i
1	9	1	4
2	8	0	5
3	6	1	7
4	10	1	3

Now let's compile the pieces that go into the partial likelihood contributions at each failure time:

ordered failure time X_i	$\mathcal{R}(X_i)$	i	Likelihood contribution $\left[e^{\beta Z_i} / \sum_{j \in \mathcal{R}(X_i)} e^{\beta Z_j}\right]^{\delta_i}$
6	{1,2,3,4}	3	$e^{7\beta}/[e^{4\beta} + e^{5\beta} + e^{7\beta} + e^{3\beta}]$
8	{1,2,4}	2	1
9	{1,4}	1	$e^{4\beta}/[e^{4\beta} + e^{3\beta}]$
10	{4}	4	$e^{3\beta}/e^{3\beta} = 1$

The partial likelihood would be the product of these four terms.

2. Likelihood

A censored-data likelihood derivation:

Recall that in general, the likelihood contributions for right-censored data fall into two categories:

- **Individual is censored at X_i:**

$$L_i(\boldsymbol{\beta}) = S_i(X_i)$$

- **Individual fails at X_i:**

$$L_i(\boldsymbol{\beta}) = f_i(X_i) = S_i(X_i)\lambda_i(X_i)$$

Thus, for iid data the full likelihood is

$$\begin{aligned} L(\boldsymbol{\beta}) &= \Pi_{i=1}^{n} L_i(\boldsymbol{\beta}) = \Pi_{i=1}^{n}\{ [f_i(x_i)]^{\delta i} [S_i(x_i)]^{1-\delta i} \} \\ &= \Pi_{i=1}^{n}\{ [S_i(x_i)\, \lambda_i(x_i)]^{\delta i} [S_i(x_i)]^{1-\delta i} \} \\ &= \Pi_{i=1}^{n}\{ [\lambda_i(x_i)]^{\delta i} [S_i(x_i)]^{1} \} \end{aligned}$$

So the full likelihood is:

$$\begin{aligned} L(\boldsymbol{\beta}) &= \prod_{i=1}^{n} \lambda_i(X_i)^{\delta_i} S_i(X_i) \\ &= \prod_{i=1}^{n} \left[\frac{\lambda_i(X_i)}{\sum_{j \in \mathcal{R}(X_i)} \lambda_j(X_i)} \right]^{\delta_i} \left[\sum_{j \in \mathcal{R}(X_i)} \lambda_j(X_i) \right]^{\delta_i} S_i(X_i) \end{aligned}$$

in the above we have multiplied and divided by the term $\left[\sum_{j \in \mathcal{R}(X_i)} \lambda_j(X_i) \right]^{\delta_i}$.

Cox (1972) argued that the first term in this product contained almost all of the information about $\boldsymbol{\beta}$, while the last two terms contained the information about $\lambda_0(t)$, the baseline hazard.

If we keep only the first term, then under the PH assumption:

$$\begin{aligned}
L(\boldsymbol{\beta}) &= \prod_{i=1}^{n}\left[\frac{\lambda_i(X_i)}{\sum_{j\in\mathcal{R}(X_i)}\lambda_j(X_i)}\right]^{\delta_i} \\
&= \prod_{i=1}^{n}\left[\frac{\lambda_0(X_i)\exp(\boldsymbol{\beta}'\mathbf{Z}_i)}{\sum_{j\in\mathcal{R}(X_i)}\lambda_0(X_i)\exp(\boldsymbol{\beta}'\mathbf{Z}_j)}\right]^{\delta_i} \\
&= \prod_{i=1}^{n}\left[\frac{\exp(\boldsymbol{\beta}'\mathbf{Z}_i)}{\sum_{j\in\mathcal{R}(X_i)}\exp(\boldsymbol{\beta}'\mathbf{Z}_j)}\right]^{\delta_i}
\end{aligned}$$

This is the partial likelihood defined by Cox. Note that it does not depend on the underlying hazard function $\lambda_0(\cdot)$.

Note that the "PH Assumption" (Cox, 1972) is

$$\lambda(t|\mathbf{Z}) = \lambda_0(t)\exp(\boldsymbol{\beta}'\mathbf{Z})$$

That is,

$$\lambda(t|\mathbf{Z}) = \lambda_0(t)\exp(\beta_1 Z_1 + \beta_2 Z_2 + \cdots + \beta_p Z_p)$$

with

$$\begin{aligned}
\lambda(t|\mathbf{Z}=0) &= \lambda_0(t)\exp(\beta_1 * 0 + \beta_2 * 0 + \cdots + \beta_p * 0) \\
&= \lambda_0(t)
\end{aligned}$$

which is the "baseline hazard function".

3. Property of the PH-Regression (Large-Sample Results):

Cox recommended to treat the partial likelihood as a regular likelihood for making inferences about β, in the presence of the nuisance parameter $\lambda_0(\cdot)$. This turned out to be valid (Tsiatis 1981, Andersen and Gill 1982, Murphy and van der Vaart 2000).

The **log-partial likelihood** is:

$$
\begin{aligned}
\ell(\boldsymbol{\beta}) &= \log \left[\prod_{i=1}^{n} \frac{e^{\boldsymbol{\beta}'\mathbf{Z}_i}}{\sum_{\ell \in \mathcal{R}(X_i)} e^{\boldsymbol{\beta}'\mathbf{Z}_\ell}} \right]^{\delta_i} \\
&= \sum_{i=1}^{n} \delta_i \left[\boldsymbol{\beta}'\mathbf{Z}_i - \log \left\{ \sum_{\ell \in \mathcal{R}(X_i)} e^{\boldsymbol{\beta}'\mathbf{Z}_\ell} \right\} \right] \\
&= \sum_{i=1}^{n} l_i(\boldsymbol{\beta})
\end{aligned}
$$

where l_i is the log-partial likelihood contribution from individual i.

Note that the l_i's are not i.i.d. terms (why, and what is the implication of this fact?).

The **partial likelihood score function** is:

$$
U(\beta) = \frac{\partial}{\partial \beta} \ell(\beta) = \sum_{i=1}^{n} \delta_i \left\{ Z_i - \frac{\sum_{\ell \in \mathcal{R}(X_i)} Z_\ell e^{\beta' Z_\ell}}{\sum_{\ell \in \mathcal{R}(X_i)} e^{\beta' Z_\ell}} \right\}
$$

The maximum partial likelihood estimator can be found by solving $U(\beta) = 0$.

Analogous to standard likelihood theory, it can be shown that asymptotically

$$\frac{(\hat{\beta}-\beta)}{\mathrm{se}(\hat{\beta})} \sim N(0,1).$$

The variance of $\hat{\beta}$ can be estimated by inverting the second derivative of the partial likelihood,

$$\widehat{\mathrm{Var}}(\hat{\beta}) = \left[-\frac{\partial^2}{\partial\beta^2}\ell(\hat{\beta})\right]^{-1}.$$

Confidence intervals for the <u>Hazard Ratio</u>:

Many software packages provide estimates of β, but the hazard ratio, or relative risk, RR= $\exp(\beta)$ is often the parameter of interest.

Confidence intervals for $\exp(\beta)$

Form a confidence interval for $\hat{\beta}$, and then exponentiate the endpoints:

$$[L, U] \;=\; [e^{\hat{\beta}-1.96se(\hat{\beta})}, e^{\hat{\beta}+1.96se(\hat{\beta})}]$$

Hypothesis Tests:

For each covariate of interest, the null hypothesis is

$$H_0 : RR_j = 1 \Leftrightarrow \beta_j = 0$$

A Wald test of the above hypothesis is constructed as:

$$Z = \frac{\hat{\beta}_j}{\text{se}(\hat{\beta}_j)} \quad \text{or} \quad \chi^2 = \left(\frac{\hat{\beta}_j}{\text{se}(\hat{\beta}_j)}\right)^2$$

Note: if we have a factor A with a levels, then we would need to construct a χ^2 test with $(a-1)$ df, using a test statistic based on a quadratic form:

$$\chi^2_{(a-1)} = \widehat{\boldsymbol{\beta}}'_A \text{Var}(\widehat{\boldsymbol{\beta}}_A)^{-1} \widehat{\boldsymbol{\beta}}_A$$

where $\boldsymbol{\beta}_A = (\beta_1, ..., \beta_{a-1})'$ are the $(a-1)$ coefficients corresponding to the binary variables $Z_1, ..., Z_{a-1}$.

Likelihood Ratio Test:

Suppose there are $(p+q)$ explanatory variables measured:

$$Z_1, \ldots, Z_p, Z_{p+1}, \ldots, Z_{p+q}.$$

Consider the following models:

- **Model 1:** (contains only the first p covariates)

$$\lambda(t|\mathbf{Z}) = \lambda_0(t) \exp(\beta_1 Z_1 + \cdots + \beta_p Z_p)$$

- **Model 2:** (contains all $(p+q)$ covariates)

$$\lambda_i(t|\mathbf{Z}) = \lambda_0(t) \exp(\beta_1 Z_1 + \cdots + \beta_{p+q} Z_{p+q})$$

We can construct a **likelihood ratio** test for testing

$$H_0 : \beta_{p+1} = \cdots = \beta_{p+q} = 0$$

as:

$$\chi^2_{LR} = -2 \left\{ \log L(M1) - \log L(M2) \right\},$$

where $L(M\cdot)$ is the maximized partial likelihood under each model. Under H_0, this test statistic is approximately distributed as χ^2 with q df.

See the file "PH-Regression_Likelihood" for a real-life ***example*** in the end of the file.