How to handle ties with a Cox model

Example with 5 subjects

id	time (t)	status (δ)	Z
1	t_1	1	Z_1
2	t_2	1	Z_2
3	t_3	0	Z_3
4	t_4	1	Z_4
5	t_5	1	Z_5

Assume:

$t_1 = t_2 < t_3 < t_4 < t_5$

subject 1 and 2 have <u>the same</u> time to event

Cox model

$$h(t|Z_i) = h_0(t)\exp(\beta Z_i)$$

Partial likelihood:

$$L(\beta) = \prod_{i=1}^{D} \frac{\exp\left(\sum_{k=1}^{p} \beta_k Z_{(i)k}\right)}{\sum_{j\in R(t_i)} \exp\left(\sum_{k=1}^{p} \beta_k Z_{jk}\right)}$$

where:

D: set of individuals who experienced the event of interest

t_i : observed time for individual "i" (in set D)

$Z_{(i)k}$: k^{th} covariate for individual with failure time t_i

$R(t_i)$: risk set at time t_i

The partial likelihood can also be written as

$$L(\beta) = \prod_{j=1}^{n} \left(\frac{\exp(\beta^T Z_{(j)})}{\sum_{i \in R(t_j)} \exp(\beta^T Z_i)} \right)^{\delta_j}$$

where

$$\delta_i = \begin{cases} 0 & \text{if } t_j \text{ is a censoring time} \\ 1 & \text{otherwise} \end{cases}$$

Back to the example:

We have 3 different survival times: t_1, t_4, t_5. The partial likelihood is as follows

$$L(\beta) = L_1(\beta)\, L_2(\beta)\, L_3(\beta)$$

In general $L_j(\beta)$ is the element in the partial likelihood that corresponds to the j^{th} distinct survival time

$$L_2(\beta) = \frac{e^{\beta z_4}}{e^{\beta z_4} + e^{\beta z_5}}$$

There are different ways to deal with ties. Here we look at two different ways.

Note that the partial likelihood assumes that there are <u>no ties</u>.

The order of the event does matter
Each subject who experiences the event has his own contribution to the partial likelihood function

$$\sum_{j \in R(t_i)}$$

sum for all subjects at risk at the moment in which the event for the particular subject is observed

Suppose there are two subjects "m" and "n" experiencing the event at the same time.
How should we proceed?
Shall we consider subject "m" at risk while "n" experience the event or should we do the other way around?

In order to consider ties in the Cox model we have to adjust the partial likelihood

We consider here two methods:

1) Exact method:

without any knowledge about the true ordering of the survival times we have to consider all possible ordering : $2! = 2$
This method assume that the survival time of individual 1 and individual 2 (in the example) are different

Two options: consider A_1 and A_2 events:

A_1: patient 1 died before patient 2 (or subject 1 experienced the event of interest before subject 2 $t_1 < t_2$)

A_2: $t_2 < t_1$ (subject 2 experienced the event before individual 1)

Then

$L_1(\beta) =$ P(observe two deaths at time t_1)

$$P(A_1) = \frac{e^{x_1\beta}}{\sum_{l=1}^{5} e^{x_l\beta}} \times \frac{e^{x_2\beta}}{\sum_{l=2}^{5} e^{x_l\beta}}$$

$$P(A_2) = \frac{e^{x_2\beta}}{\sum_{l=1}^{5} e^{x_l\beta}} \times \frac{e^{x_1\beta}}{\sum_{\substack{l=1,\\ l\neq 2}}^{5} e^{x_l\beta}}$$

N.B! the risk set

The risk set changes according to the subject who experiences the event of interest as first one

As illustrated before

$$L(\beta) = L_1(\beta) \cdot L_2(\beta) \cdot L_3(\beta)$$

When $L(\beta)$ is known, inference is as usual. Maximize $L(\beta)$ to find $\hat{\beta}$

This method is computationally very intensive. For d_j tied survival times there are $d_j!$ different orderings to be considered and $L_j(\beta)$ is then the sum of $d_j!$ terms

2) Breslow's approximation

$$\frac{e^{\beta z_2}}{\sum_{l=2}^{5} e^{\beta z_l}} \approx \frac{e^{\beta z_2}}{\sum_{l=1}^{5} e^{\beta z_l}}$$

$$e^{\beta z_2} + e^{\beta z_3} + e^{\beta z_4} e^{\beta z_5} \approx e^{\beta z_1} + e^{\beta z_2} + e^{\beta z_3} + e^{\beta z_4} + e^{\beta z_5}$$

$$\frac{e^{\beta z_1}}{\sum_{\substack{l=1,\\ l\neq 2}}^{5} e^{\beta z_l}} \approx \frac{e^{\beta z_1}}{\sum_{l=1}^{5} e^{\beta z_l}}$$

$$e^{\beta z_1} + e^{\beta z_3} + e^{\beta z_4} + e^{\beta z_5} \approx e^{\beta z_1} + e^{\beta z_2} + e^{\beta z_3} + e^{\beta z_4} + e^{\beta z_5}$$

We can therefore approximate $P(A_1)$ and $P(A_2)$ as

$$\frac{e^{z_1\beta}}{\sum_{l=1}^{5} e^{z_l\beta}} \times \frac{e^{z_2\beta}}{\sum_{l=1}^{5} e^{z_l\beta}} = \frac{e^{(z_1+z_2)\beta}}{\left(\sum_{l=1}^{5} e^{z_l\beta}\right)^2}$$

If we have d_j tied survival times the likelihood $L_j(\beta)$ at the j^{th} distinct survival times is approximate as

$$L_j(\beta) \approx \frac{\exp\left(\beta \sum_{l\in D_j} z_l\right)}{\left[\sum_{l\in R_j} \exp(z_l\beta)\right]^{d_j}}$$