Kaplan Meier Analysis

Hila Shacham

February 8, 2021

1 Logrank test

- t - Number of (unique) events.
- g - Number of groups.
- $O - [t \times g]$ Matrix of observed events (Each row represents an event, while each column represents a group).
- $R - [t \times g]$ Matrix of risk groups.
- $E - [t \times g]$ Matrix of expected values.

At event i for $i \in \{2, .., t\}$:

$$R_i = R_{i-1} - O_{i-1} - C_{i-1} \tag{1}$$

Where C_{i-1} is the number of population that was censored at event i-1. C is not a defined matrix in the matlab calculations, C_{i-1} was only added for the sake of clarification of this document. The censorship is brought into account in the observed O matrix.

Initialized values:

- R_1 - The total number of population at the beginning.

Other matrices:

- $R_{sum} - [t \times 1]$ Sum of the risk values over all the groups for different time events. R_{sum_i} is the sum of risk values at time/event i over all the groups.
- $O_{sum} - [t \times 1]$ Sum of the observed values over all the groups for different time events. O_{sum_i} is the sum of observed events at time/event i over all the groups.

For time i and group j:

$$E_{ij} = \frac{R_{ij} \cdot O_{sum_i}}{R_{sum_i}} \tag{2}$$

More matrices:

- $Var - [t \times (g-1)]$ Variance matrix.
- $Var_{sum} - [1 \times (g-1)]$ The sum of each group's variance over all the time events.
- $COV - [t \times \frac{(g-1)\cdot(g-2)}{2}]$ The Covariance matrix. As for the columns' size: $\frac{(g-1)\cdot(g-2)}{2} = \binom{g-1}{2}$, i.e. the number of all possible pairs of a group of size $|g-1|$.

For time i and group j where $j \in \{1, 2, ..., g-1\}, i \in \{1, ..., t\}$:

$$Var_{ij} = \frac{R_{ij} \cdot (R_{sum_i} - R_{ij}) \cdot O_{sum_i} \cdot (R_{sum_i} - O_{sum_i})}{R_{sum_i}^2 \cdot (R_{sum_i} - 1)} \tag{3}$$

Pay attention that we only take all the pairs of the first (g-1) groups, all while excluding elements from the last group. I have no idea why though???
Same happens with the covariance:

Algorithm 1 Calculating Covariance

```
for time k ∈ 1, ..., t do
    l = 0
    for i ∈ {1, ..., g − 2} do
        for j ∈ {i, ..., g − 1} do
            l = l + 1
            COV_kl = (−R_ki · R_kj · O_sum_k · (R_sum_k − O_sum_k)) / (R_sum_k^2 · (R_sum_k − 1))
```

$$COV_{kl} = \frac{-R_{ki} \cdot R_{kj} \cdot O_{sum_k} \cdot (R_{sum_k} - O_{sum_k})}{R_{sum_k}^2 \cdot (R_{sum_k} - 1)}$$

- $Var_{sum} - [1 \times (g-1)]$ Sum of all the values in matrix V for each group over the time events.
- V - $[(g-1) \times (g-1)]$ Square matrix containing elements from the variance matrix Var and the covariance matrix COV.

The upper and lower triangulars of V_{sum} are elements of COV, while the diagonal consists of Var. $\frac{(g-1)\cdot(g-2)}{2}$ is indeed the size of each triangular, while (g-1) is the size of the diagonal.

$$\text{V} = \begin{bmatrix} \ddots & & COV \\ & Var_{sum} & \\ COV & & \ddots \end{bmatrix}$$

We also calculate:
$d - [(g-1) \times 1]$
For group j, where $j \in \{1, .., g-1\}$

$$d_{j1} = \sum_{i=1}^{t} O_{ij} - E_{ij} \tag{4}$$

And now we can calculate χ^2:

$$\chi^2 = \frac{d^\dagger}{V} \cdot d \tag{5}$$