

Variable Selection For Discrete Competing Risks Models

Jingwei Zhang

skykiny@outlook.com

June 28, 2016

Statistical Model

The Discrete Competing Risk Model

- Consider an example involving multiple causes of failure.
- For a patient who took surgery, we want to study the occurrence of different types of events, such as disease recurrence and death.
- This patient is followed up for several times. In every follow-up, we care about whether these event happens or not.
 - If **one** of these events happens, the follow up will **stop**.
 - If **none** of these event happens(usually we call this situation **survival**), we **continue** to monitor this patient.
 - After several times of follow up, the monitor will stop(we call it **censoring**).

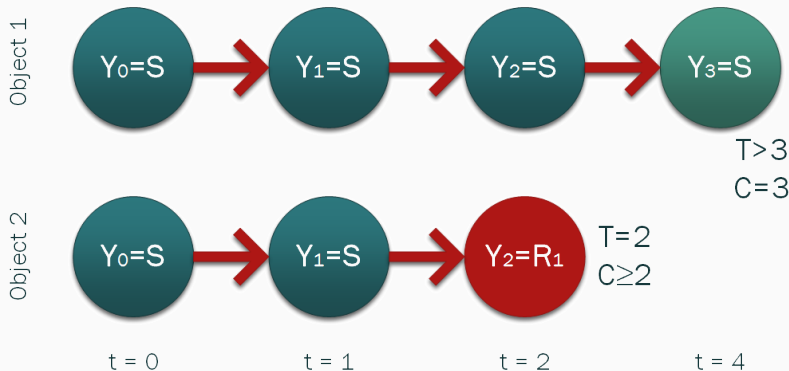
Purpose of the Model

This model can be used to:

- study the relationship between a vector of covariates \mathbf{x} and the rate of occurrence of specific types of failure.
- estimate the risk of one type of failure after removing others.

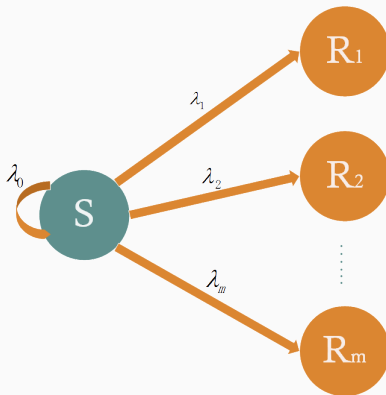
Example of State Transition

- $T \in \{1, \dots, k\}$ indicates the **first time** of happening of event, C indicates the time of censoring.
- Y_t indicates the status of the object on time t . R_1, \dots, R_m indicates m events and S indicates survival.



Graph of State Transition

- $\lambda_r(t|\mathbf{x})$ represents the transitional probability from state S to state R_r . $r = 0$ represents no event happens.
- We call $\lambda_r(t|\mathbf{x})$ cause-specific discrete hazard function.
- Both \mathbf{x} and t affects probability of transition.



Hazard Functions

The **cause-specific hazard function**:

$$\begin{aligned}\lambda_r(t|\mathbf{x}) &= P(Y_t = R_r | Y_{t-1} = S, \mathbf{x}) \\ &= P(T = t, R = r | T \geq t, \mathbf{x})\end{aligned}$$

Summing up for all events, we get **hazard function**:

$$\lambda(t|\mathbf{x}) = \sum_{r=1}^m \lambda_r(t|\mathbf{x}) = P(T = t | T \geq t, \mathbf{x})$$

For survival, we have:

$$\lambda_0(t|\mathbf{x}) = 1 - \lambda(t|\mathbf{x}) = 1 - \sum_{r=1}^m \lambda_r(t|\mathbf{x})$$

Unconditional Probability

The **survival function**, which indicates the unconditional probability of no event happening:

$$S(t|\mathbf{x}) = P(T > t|\mathbf{x}) = \prod_{j=1}^t (1 - \lambda(j|\mathbf{x}))$$

The unconditional probability of one event happening on time t :

$$\begin{aligned} P(T = t|\mathbf{x}) &= \lambda(t|\mathbf{x}) \prod_{j=1}^{t-1} (1 - \lambda(j|\mathbf{x})) \\ &= \lambda(t|\mathbf{x}) S(t-1|\mathbf{x}) \end{aligned}$$

The Multinomial Logit Model

To model the influence of \mathbf{x} on transitional probability λ_r , multinomial logit model is used:

$$\lambda_r(t|\mathbf{x}) = \frac{\exp(\beta_{0tr} + \mathbf{x}^T \gamma_r)}{1 + \sum_{s=1}^m \exp(\beta_{0ts} + \mathbf{x}^T \gamma_s)}$$

where $t = 1, \dots, q$, and $r = 1, \dots, m$.

- $\beta_{01r}, \dots, \beta_{0qr}$ determine the cause-specific baseline hazard functions.
- γ_r contains the cause-specific effects of covariates.

For survival, we have:

$$\lambda_0(t|\mathbf{x}) = 1 - \sum_{r=1}^m \lambda_r(t|\mathbf{x}) = \frac{1}{1 + \sum_{s=1}^m \exp(\beta_{0ts} + \mathbf{x}^T \gamma_s)}$$

Original Data

Data is given by $(t_i, r_i, \delta_i, \mathbf{x}_i)$, $i = 1, \dots, n$

- $t_i = \min(T_i, C_i)$ is the observed discrete time, which is the minimum of survival time T_i and censoring time C_i .
- We always assume **random censoring**, that is, T_i and C_i are assumed to be independent
- $r_i \in \{1, \dots, m\}$ indicates the type of the terminating event
- \mathbf{x}_i is the covariate vector
- δ_i denotes the censoring indicator with

$$\delta_i = \begin{cases} 1, & \text{if event occurred on time } t_i \\ 0, & \text{if it censors on time } t_i \end{cases}$$

Likelihood Function

The likelihood contribution of the i -th observation is given by (\mathbf{x} is omitted):

$$L_i = P(T_i = t_i, R_i = r_i)^{\delta_i} P(T_i > t_i)^{1-\delta_i} P(C_i \geq t_i)^{\delta_i} P(C_i = t_i)^{1-\delta_i}$$

Under the assumption that censoring does not depend on the parameters that determine the survival time (**non-informative censoring**), L_i can be reduced to

$$\begin{aligned} L_i &= P(T_i = t_i, R_i = r_i | \mathbf{x})^{\delta_i} P(T_i > t_i | \mathbf{x})^{1-\delta_i} \\ &= \lambda_{r_i}(t_i | \mathbf{x}_i)^{\delta_i} (1 - \lambda(t_i | \mathbf{x}_i))^{1-\delta_i} \prod_{t=1}^{t_i-1} (1 - \lambda(t | \mathbf{x}_i)) \end{aligned}$$

Response Vector

For an alternative form of the likelihood, indicators for the transition to the next period are defined by

$$y_{itr} = \begin{cases} 1, & \text{if event of type } r \text{ occurred on time } t \\ 0, & \text{if event of type } r \text{ did not occur on time } t \end{cases} \quad (1)$$

and

$$y_{it0} = \begin{cases} 1, & \text{if no events occurred on time } t \text{ (survive)} \\ 0, & \text{if one of the } m \text{ events occurs on time } t \end{cases} \quad (2)$$

where $i \in R_t$ and $r = 1, \dots, m$. These indicator variables are gathered in the vector $\mathbf{y}_{it}^T = (y_{it0}, y_{it1}, \dots, y_{itm})$ denoting the response vector of object i , $i = 1, \dots, n$, $t = 1, \dots, t_i$.

Response Vector - Example

$$\mathbf{y}_{it} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

where $y_{it0} = 1$ while others are 0. This vector indicates that on time t , the object i survives.

$$\mathbf{y}_{it} = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}$$

where $y_{itj} = 1$ while others are 0. This vector indicates that on time t , event j happens.

Rewrite the Likelihood Function

Using response vector, the likelihood function L_i can be rewritten as:

$$L_i = \prod_{t=1}^{t_i} \left(\prod_{r=0}^m \lambda_r(t|\mathbf{x}_i)^{y_{itr}} \right)$$

This is actually a multinomially distribution with

$$\mathbf{y}_{it}^\top = (y_{it0}, y_{it1}, \dots, y_{itm}) \sim \mathcal{M}(1, (\lambda_0(t|\mathbf{x}), \lambda_1(t|\mathbf{x}), \dots, \lambda_m(t|\mathbf{x})))$$

Log Likelihood Function

The total log-likelihood is given by

$$l = \sum_{i=1}^n \sum_{t=1}^{t_i} \sum_{r=0}^m y_{itr} \log \lambda_r(t|\mathbf{x})$$

where $\lambda_r(t|\mathbf{x})$ is given by multinomial logit model:

$$\lambda_r(t|\mathbf{x}) = \frac{\exp(\eta_{itr})}{1 + \sum_{s=1}^m \exp(\eta_{its})}$$

where $\eta_{itr} = \beta_{0tr} + \mathbf{x}^T \gamma_r$.

This ML estimates can be easily computed by using statistical software for multinomial regression models.