PH 716 Applied Survival Analysis

Part 10: Recurent Events

Zhiyang Zhou (zhou67@uwm.edu, zhiyanggeezhou.github.io)

2026/05/07 15:56:08

Motivation

- Standard survival analysis
  - Focus: time to first event
  - Example: time to death, first relapse
- Recurrent event setting
  - Individuals can experience multiple events over time
  - Examples: hospital readmissions, asthma attacks, tumor recurrences, equipment breakdowns

A concrete example: `survival::bladder1`

Full dataset from a study on recurrences of bladder cancer. It contains all three treatment arms and all recurrences for 118 subjects; the maximum number of recurrences is 9.

- `id`: Patient id
- `treatment`: Placebo, pyridoxine (vitamin B6), or thiotepa
- `number`: Initial number of tumours
- `size`: size (cm) of largest initial tumour
- `recur`: Number of recurrences
- `start`, `stop`: The start and end time of each time interval
- `status`: End of interval code (`0=censored`, `1=recurrence`, `2=death from bladder disease`, `3=death other/unknown cause`)
- `rtumor`: Number of tumors found at the time of a recurrence
- `rsize`: Size of largest tumor at a recurrence
- `enum`: Event number within patient

```
head(survival::bladder1)
```

```
##   id treatment number size recur start stop status rtumor rsize enum
## 1  1   placebo      1    1     0     0    0      3      .     .    1
## 2  2   placebo      1    3     0     0    1      3      .     .    1
## 3  3   placebo      2    1     0     0    4      0      .     .    1
## 4  4   placebo      1    1     0     0    7      0      .     .    1
## 5  5   placebo      5    1     0     0   10      3      .     .    1
## 6  6   placebo      4    1     1     0    6      1      1     1    1
```

Why standard Cox model is insufficient

The usual Cox model assumes:

- Only one event per subject
- Independence across observations

But for recurrent events:

- Multiple events per subject
- Within-subject dependence
- Risk may change after each event

Notations for recurrent events

- T_{ik}: authentic time of the kth event for subject i
- $N_i(t)$: number of events that have happened up to time t for subject i, forming a **counting process**
- $Y_i(t)$: time-varying at-risk indicator, indicating whether subject i is under observation just before time t
- x_{ij}: the jth covariate for subject i

Counting process

Imagine each person carries a counter:

- Every time an event happens -> the counter goes up by 1
- Between events -> the counter stays the same

That counter is a counting process:

- In time-to-first-event: the counter can jump at most once.
- In recurrent events: it looks like a step function: flat -> jump -> flat -> jump -> flat -> jump

We track:

- For time-to-first-event: "How fast does the first jump happen?" (i.e., the hazard function)
- For recurrent events: "How fast do the jumps happen over time?" (i.e., the **intensity function**, an extension of the hazard function)

Intensity function

- Define the increment over a time interval $[t, t + \mathrm{d}t)$, $\mathrm{d}t > 0$

  $$\mathrm{d}N_i(t) = N_i(t + \mathrm{d}t) - N_i(t)$$

  Then $\mathrm{d}N_i(t)$ is an indicator of whether an event happens in $[t, t + \mathrm{d}t)$ (with $\mathrm{d}t$ so small that there is at most one event within $[t, t + \mathrm{d}t)$):

  - $\mathrm{d}N_i(t) = 1$ if an event happens in $[t, t + \mathrm{d}t)$
  - $\mathrm{d}N_i(t) = 0$ if no event happens in $[t, t + \mathrm{d}t)$
  - Reducing to the usual failure indicator in time-to-first-event setting

- The counting process $N_i(t)$ is characterized via its intensity:

  $$\lambda_i(t) = \lim_{\mathrm{d}t \to 0} \frac{\Pr(\mathrm{d}N_i(t) = 1 \mid \mathcal{F}_{t-})}{\mathrm{d}t}$$

  - $\mathcal{F}_{t-}$: the history of events, censoring, and covariates right before time t
  - Interpretation: the instantaneous event rate at time t, given the past history of events and covariates
  - Extension of the hazard function to recurrent events

- Expected cumulative number of events by time t

  $$\mathrm{E}\{N_i(t)\} = \int_0^t \lambda_i(u)\mathrm{d}u$$

- Probability of no recurrent event occurring by time t

  $$\Pr(N_i(t) = 0) = \exp\left(-\int_0^t \lambda_i(u)\mathrm{d}u\right)$$

Andersen-Gill model

- Assume:
  - Recurrent events as repeated jumps of the same counting process
  - No within-subject dependence
- Andersen-Gill model

$$\lambda_i(t) = Y_i(t)\lambda_0(t)\exp\left(\sum_{j=1}^{p} x_{ij}\beta_j\right)$$

  - $\lambda_0(t)$: baseline intensity function
  - β_j: regression coefficient for covariate j

Ex. 10.1 Revisit `survival::bladder1`

Only recurrences (`status==1`) count as events in the counting process. Death stops observation, acting like terminal censoring.

```r
options(digits=4)
library(survival)
df = subset(survival::bladder1, stop > start)
df$event = ifelse(df$status==1, 1, 0) # terminal events taken as censoring
df$Treatment = factor(df$treatment)
df$number = factor(df$number)
df$rsize = factor(df$rsize)
# Andersen-Gill model
## NOT accounting for any within-subject dependence
fit_ag_1 = coxph(
  Surv(start, stop, event) ~ Treatment + number + rsize,
  data = df
)
summary(fit_ag_1)
## Accounting for within-subject dependence (via robust variance estimation)
fit_ag_2 = coxph(
  Surv(start, stop, event) ~ Treatment + number + rsize + cluster(id),
  data = df
)
summary(fit_ag_2)

# Expected cumulative number of events up to time 20 for a new subject
base = basehaz(fit_ag_2, centered = FALSE) # baseline cumulative intensity function
newx = data.frame(
  Treatment = factor("thiotepa", levels = levels(df$Treatment)),
  number = factor(2, levels = levels(df$number)),
  rsize = factor(2, levels = levels(df$rsize))
)
lp = predict(fit_ag_2, newdata = newx, type = "lp")
(expected_num_events = exp(lp) *  max(base$hazard[base$time <= 20]))

# Further give the probability of no recurrent event occurring by time 20 for the same subject
(prob_no_event = exp(-expected_num_events))
```