Statistical Practice in Epidemiology 2018

Survival analysis with competing risks

Janne Pitkäniemi (EL)

Points to be covered

1. Survival or time to event data & censoring.
2. Competing risks: event-specific cumulative incidences & hazards.
3. Kaplan–Meier and Aalen–Johansen estimators.
4. Regression modelling of hazards: Cox model.
5. Packages `survival`, `mstate`, `cmprisk`.
6. Functions `Surv()`, `survfit()`, `plot.survfit()`, `coxph()`, `Cuminc()`.

Survival time – time to event

Time spent (`lex.dur`) in a given **state** (`lex.Cst`) from its beginning till a certain *endpoint* or *outcome* **event** (`lex.Xst`) or *transition* occurs, changing the state to another.

Examples of such times and outcome events:

- lifetime: birth $\rightarrow$ death,
- duration of marriage: wedding $\rightarrow$ divorce,
- healthy exposure time:
  start of exposure $\rightarrow$ onset of disease,
- clinical survival time:
  diagnosis of a disease $\rightarrow$ death.

Ex. Survival of 338 oral cancer patients

Important variables:

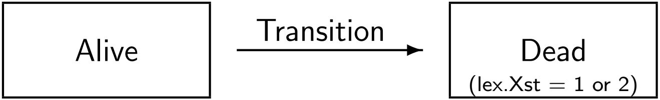
- `time` = duration of patientship from diagnosis (**entry**) till death (`death`) or censoring (`Alive`), (`lex.Cst` is (`Alive`))
- `event` = indicator for the outcome and its observation at the end of follow-up (**exit**):
  0 = censoring,
  1 = death from oral cancer

Special features:

- Two possible endpoints
- Censoring – incomplete observation of the survival time.

Set-up of classical survival analysis

- **Two-state model**: only one type of event changes the initial state.
- Major applications: analysis of lifetimes since birth and of survival times since diagnosis of a disease until death from any cause.



- **Censoring**: Death and final lifetime not observed for some subjects due to emigration or closing the follow-up while they are still alive

Distribution concepts: hazard function

The **hazard rate** or **intensity** function $\lambda(t)$

$$\lambda(t) = P(t < T \leq t + \Delta | T > t)/\Delta, \ \mathit{forsmall} \Delta$$

- $\approx$ the conditional probability that the event occurs in a short interval $(t, t + \Delta]$, given that it does not occur before t, divided by interval length.

In other words, during a short interval

$$\text{risk of event} \approx \text{hazard} \times \text{interval length}$$

Distribution concepts: survival and cumulative hazard functions

Survival function

$$S(t) = P(T > t),$$

$=$ probability of avoiding the event at least up to t
(the event occurs only after t).
The **cumulative hazard** (or integrated intensity):

$$\Lambda(t) = \int_0^t \lambda(u)du$$

Connections between the functions:

$$S(t) \;\; = \;\; \exp\{-\Lambda(t)\}$$

Observed data on survival times

For individuals $i = 1, \ldots, n$ let
T_i = time to outcome event,
U_i = time to censoring.

Censoring is assumed **noninformative**, *i.e.* independent from occurrence of events.

We observe

$y_i = \min\{T_i, U_i\}$, *i.e.* the exit time, and

$\delta_i = 1_{\{T_i < U_i\}}$, indicator $(1/0)$ for the outcome event occurring first, before censoring.

Censoring must properly be taken into account in the statistical analysis.

Approaches for analysing survival time

- **Parametric model** (like Weibull, gamma, etc.) on hazard rate $\lambda(t) \rightarrow$ Likelihood:

$$L = \prod_{i=1}^{n} \lambda(y_i)^{\delta_i} S(y_i)$$

- **Piecewise constant rate** model on $\lambda(t)$ – see Bendix's lecture on time-splitting (Poisson likelihood).

- **Non-parametric** methods, like Kaplan–Meier (KM) estimator of survival curve $S(t)$ and Cox proportional hazards model on $\lambda(t)$.

R package `survival`

Tools for analysis with one outcome event.

- `Surv(time, event) -> sobj`
  creates a **survival object** `sobj` assuming that all start at 0, containing pairs (y_i, δ_i),
- `Surv(entry, exit, event) -> sobj2`
  creates a survival object from `entry` and `exit` times,
- `survfit(sobj ~ x) -> sfo`
  creates a **survfit** object `sfo` containing KM or other non-parametric estimates (also from a fitted Cox model),
- `plot(sfo)`
  plot method for survival curves and related graphs,
- `coxph(sobj ~ x1 + x2)`
  fits a Cox model with covariates `x1` and `x2`.
- `survreg()` – parametric survival models.

Ex. Oral cancer data (cont'd)

```
> orca$suob <- Surv(orca$time, 1*(orca$event > 0) )

> orca$suob[1:7]    #  + indicates censored observation
[1] 5.081+ 0.419  7.915  2.480  2.500  0.167  5.925+

> km1 <- survfit( suob ~ 1, data = orca)
> km1                    #  brief  summary
records   n.max n.start  events  median 0.95LCL 0.95UCL
 338.00  338.00  338.00  229.00    5.42    4.33    6.92

> summary(km1)      #  detailed KM-estimate

  time n.risk n.event survival std.err lower 95% CI upper 95% CI
 0.085    338       2   0.9941 0.00417       0.9859        1.000
 0.162    336       2   0.9882 0.00588       0.9767        1.000
 0.167    334       4   0.9763 0.00827       0.9603        0.993
 0.170    330       2   0.9704 0.00922       0.9525        0.989
 0.246    328       1   0.9675 0.00965       0.9487        0.987
  ...
```

Ex. Oral cancer KM estimates