Competing Risks

Matthew Austin

Institute for Advanced Analytics
MSA Class of 2019

Introduction to competing risks:
Overview

Multiple event types

- All of the methods we've used in class so far have been for studying the time until one event occurs
- Everything we've learned can be extended to multiple events or multiple *types* of events
- Today, we'll discuss the latter: what happens when there is more than one possible reason for failure?
- All of the methods you have and will learn in this course can be generalized as a particular type of *multi-state* analysis

Competing risks

- Consider the following examples:
  - Death from cancer vs. death from other causes
  - Supreme Court justices leaving due to death vs. retirement
  - Leaving a job due to retirement, injury, or being fired
  - Pump failure due to jamming, flooding, motor failure, or surge
- In all of these cases, there are multiple, **mutually exclusive** causes of failure
- These are examples of a **competing risks** problem, where each subject can experience only one of several possible events

Previous assumptions

- "Regardless, we assume [event time] and [censoring time] are independent"
- "Basically, we act as if subjects censored at time t were randomly selected to be censored from all subjects still at risk at time t"
- "Under independence … there's actually no [mathematical] difference between fixed and random censoring"

What "independence" means

- On your HW, you focused on flooding failures and treated all other failure types as censored
- In this context, the independence assumption means that a censored pump and an uncensored pump have the same risk of flooding, *regardless of the reason for censoring*
- If censoring weren't independent, then a censored pump (say, that jammed) would a different risk of flooding than the uncensored pumps—the ones that actually flooded

Independence

- By treating other failure types as censored, we're essentially implying that once a pump fails due to jamming, we still don't know when it would fail due to flooding—we assume that the event types are independent
- There is, of course, no test for this
- You can decide whether or not independence is reasonable; i.e., if subjects with a high risk of one event are equally likely to experience the other events
- For example, you could just assume that at any time, pumps that flooded aren't any more or less likely to have jammed, had a mechanical failure, or experienced a surge of water

Introduction to competing risks:
Estimation

Data set

- In today's class, we'll model the tenure of Supreme Court justices and their reason for departure (as of 12/31/17)
- Response: how long do SCOUTS justices sit on the Court?
  - `days` is the tenure (in days, but we'll use years) on the Court, beginning on the day that the judicial oath was taken
  - `status` indicates the reason for leaving the Court: (0) currently sitting, (1) death, or (2) resignation/retirement
- Predictors:
  - `nomage`: age at nomination
  - `nomyear`: year of nomination
  - `nomchief`: indicator of whether or not the candidate was *nominated* as Chief Justice

Review

- Remember the two major functions in survival analysis: the survival function and the hazard function
- The survival function is the probability of the surviving beyond time t:

$$\begin{aligned} S(t) &= \Pr(T > t) \\ &= 1 - \Pr(T \leq t) \\ &= 1 - F(t) \end{aligned}$$

- The hazard function is the conditional failure rate in an interval:

$$\begin{aligned} \lambda(t) &= -\frac{d}{dt} \log S(t) \\ &= -\frac{d}{dt} \log(1 - F(t)) \\ &= \frac{f(t)}{1 - F(t)} \end{aligned}$$

Cause-specific hazards

- Since we are now dealing with different types of events, we need to extend what we already know to each possible event type
- For a particular event type k, the **cause-specific hazard** is the conditional failure rate *from cause* k:

$$\lambda_k(t) = \frac{f_k(t)}{1 - F(t)}$$

- Same interpretation as before, just cause-specific now; e.g., rate of death *when only death can happen*
- The overall hazard is just the sum of the cause-specific hazards:

$$\lambda(t) = \sum_k \lambda_k(t)$$

Cumulative incidence function

- Focusing again on cause-specific estimates, the **cumulative incidence function** (CIF) is the [*unconditional!*] probability that cause k occurs *by* time t:

$$F_k(t) = \Pr(T \leq t, K = k) = \int_0^t \lambda_k(u) S(u) du$$

- The probability of *any* event by time t is just the sum of the individual CIFs:

$$F(t) = \sum_k F_k(t)$$

and it should be clear that $S(t) = 1 - F(t) = 1 - \sum_k F_k(t)$

Overall survival

$$S(t) = 1 - \sum_k F_k(t)$$

- Note that the survival function is still unconditional, since survival means surviving **all** of the risks, so there's no such thing "cause-specific" survival
- Thus, because $S(t)$ will change when *any* event happens, the probability/incidence of cause k will also change—whereas it wouldn't if we just censored all non-k events (since censoring doesn't change the survival probability)

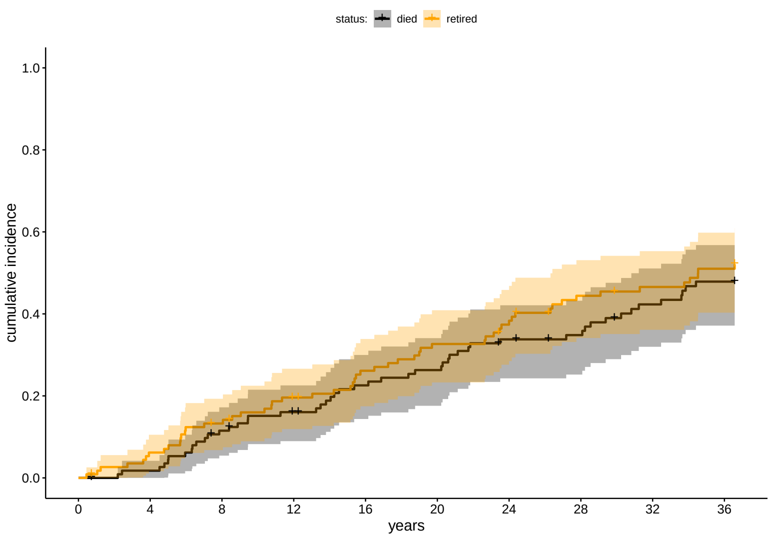
Estimating the CIF

We can estimate the CIFs using the nonparametric estimates of $\hat{S}(t), \hat{\lambda}(t)$ that we learned previously:

$$\hat{F}_k(t) = \sum_{t_m \leq t} \hat{\lambda}_k(t_m)\hat{S}(t_{m-1})$$

where the cause-specific hazard $\lambda_k(t) = \frac{d_{kt}}{r_t}$ is the number of cause-k events at time t divided by the overall risk set at time t

Estimating the CIF (con't)



Modeling competing risks:
Cause-specific hazard model

The usual Cox regression model

- A typical modeling approach for competing risks is to use separate Cox regression models for each cause, treating the other events as censored
- What this approach is actually doing is modeling the effects of predictors on the cause-specific hazard:

$$\lambda_{ik}(t) = \lambda_{0k}(t)e^{\mathbf{x}_i^\top \boldsymbol{\beta}}$$

- When doing this, we're focused on just one cause, basically ignoring anything else that can happen; e.g., we create a hypothetical world where death is the only reason Supreme Court justices ever leave the Court

Cause-specific model: death

```
cox_death <- coxph(Surv(time = days/365.25,
                        event = status == 1) ~ nomage +
                   nomyear + nomchief, data = scotus)
```

	exp(coef)	lower .95	upper .95
nomage	1.075	1.025	1.127
nomyear	0.988	0.983	0.993
nomchief	0.642	0.260	1.587

- Older justices are at a higher risk of dying in office (HR 95% CI: [1.025, 1.127])
- The risk of dying in office decreases each year (HR 95% CI: [0.983, 0.993])
- (When only death can occur)

Modeling competing risks:
Fine-Gray model

Sub-hazards

- By treating non-k events as censored, you are modeling the cause-specific hazard:

$$\lambda_k(t) = -\frac{d}{dt}(\log(1 - F(t)))$$

  which is the rate of, say, death at time t for those still at risk at time t

- Instead, let's take the CIFs and look at the **sub-hazard**:

$$\lambda_k^s(t) = -\frac{d}{dt}(\log(1 - F_k(t)))$$

  which can be thought of as the rate of death at time t conditional on **not having died yet**—so justices departing for reasons other than death *are still "allowed" to die!*

Proportional sub-hazard model

- In the **Fine-Gray model** (FG), we thus replace the cause-specific hazard with the sub-hazard to get a "proportional sub-hazard" (PSH) model:

$$\lambda^s_{ik}(t) = \lambda^s_{0k}(t)e^{\mathbf{x}_i^\top \boldsymbol{\beta}}$$

- By using the CIFs rather than overall survival (i.e., $1 - F_k(t)$ vs. $1 - F(t)$), the risk set is adjusted to account for different events happening
- In this framework, the risks are essentially balanced by looking at the *probability* of one specific cause in the presence of other causes by keeping failures from other causes "at risk" for the particular one we're focused on
- When using CIFs, we don't need to assume the competing risks are independent!

Sub-hazard model: death

```
fg_death <- coxph(Surv(fgstart, fgstop, fgstatus) ~ nomage +
                    nomyear + nomchief,
                  data = death_fgdata, weight = fgwt)
```

	exp(coef)	lower .95	upper .95
nomage	1.007	0.968	1.048
nomyear	0.991	0.986	0.996
nomchief	0.850	0.358	2.020

- Nomination age doesn't have much of an effect on the probability of dying in office (HR 95% CI: [0.968, 1.048])
- Dying in office becomes a little less likely each year (HR 95% CI: [0.986, 0.996])
- (When either death or retirement can happen)