Analysis of Survival Data

Competing Risks (2014)

1 Introduction

1.1 Context

A *competing risks* analysis is a time-to-event analysis where (1) individuals are at risk of more than one different event and (2) after the first event the individual is not at risk of a second event.

A medical example would be a clinical trial where individuals are at risk of dying from several causes.

1.2 Statistical Model

Each individual is at risk of two events of interest (A and B) and of censoring C. The restriction to *two* events of interest is merely to simplify the notation, there is no difficulty about extending the analysis to more than two events.

We imagine that there are two times-to-event – T_A, T_B – which are independent given a set of modelled baseline covariates Z. We will use J as a shorthand for the composite event 'A or B' so that $T_J = T_{A\cup B} = \min(T_A, T_B)$. We assume $T_A \neq T_B$: which means there is a well defined 'first' event $E \in \{A, B\}$ such that $E = A$ iff $T_J = T_A$.

The results will be presented generally in terms of event A, similar results hold for event B by symmetry.

2 Lecture

2.1 Cause-Specific Hazard

The *cause-specific hazard* $\mathrm{h}_A(t)$ for event A is the event rate at time t for A given no event of any sort (A or B) before time t:

$$\begin{aligned}\mathrm{h}_A(t) &= \lim_{\Delta\downarrow 0}\left\{\frac{1}{\Delta}\mathbb{P}\left[t < T_A \leq t + \Delta | t < T_J\right]\right\}\\ &= \lim_{\Delta\downarrow 0}\left\{\frac{1}{\Delta}\mathbb{P}\left[t < T_A \leq t + \Delta | t < T_A\right]\right\}\end{aligned}$$

where the second equality is a consequence of A and B being independent.

2.2 Competing risks

The hazard $\mathrm{h}_J(t)$ for the event death-from-any-cause ($A \cup B$) is given by the sum of hazards for the individual events:

$$\mathrm{h}_J(t) = \mathrm{h}_A(t) + \mathrm{h}_B(t)$$

where $\mathrm{h}_A(t)$ is defined as:

$$\mathrm{h}_A(t) = \lim_{\Delta \downarrow 0} \left\{ \frac{1}{\Delta} \mathbb{P}\left[t < T_A \leq t + \Delta | t < T_J\right] \right\},$$

and the same relationship holds for the integrated hazards:

$$\mathrm{H}_J(t) = \mathrm{H}_A(t) + \mathrm{H}_B(t).$$

The probability of no event of either sort in $[0, t]$ is given by:

$$\mathbb{P}\left[t < T_J\right] = \mathrm{F}_J(t) = \exp\left\{-\mathrm{H}_J(t)\right\}.$$

2.3 Cause Specific Density

The density $\mathrm{f}_A(t)$ of event A in the absence of event B is not the same as the density $\tilde{\mathrm{f}}_A(t)$ in the presence of competing event B because if $T_B < T_A$ there is no contribution to the density of A. The density $\tilde{\mathrm{f}}_A(t)$ is the *cause specific density* and is defined as:

$$\tilde{\mathrm{f}}_A(t) = \lim_{\Delta \downarrow 0} \left\{ \frac{1}{\Delta} \mathbb{P}\left[(t < T_J \leq t + \Delta) \& (T_J = T_A)\right] \right\}.$$

The (ordinary) density can be written:

$$\mathrm{f}_A(t) = \mathrm{h}_A(t)\mathrm{F}_A(t)$$

and by analogy the cause specific density is given by:

$$\tilde{\mathrm{f}}_A(t) = \mathrm{h}_A(t)\mathrm{F}_J(t)$$

where the substitution of F_J for F_A accounts for B occurring before A.

2.4 Cumulative Risk Function

The probability of an event occurring in $[0, t]$ and that event being A is known as the *cumulative risk function* for A, which will be denoted as $\mathrm{G}_A(t)$. The cumulative risk function is obtained by integrating the cause-specific density:

$$\begin{aligned} \mathrm{G}_A(t) &= \int_0^t \tilde{\mathrm{f}}_A(t')\mathrm{d}t' = \int_0^t \mathrm{h}_A(t')\mathrm{F}_J(t')\mathrm{d}t' \qquad (1) \\ &= \int_0^t \mathrm{h}_A(t') \exp\left\{ -\int_0^{t'} \mathrm{h}_A(t'')\mathrm{d}t'' - \int_0^{t'} \mathrm{h}_B(t'')\mathrm{d}t'' \right\} \mathrm{d}t'. \end{aligned}$$

Note that if $\mathrm{F}_J(\infty) = 0$ then $\mathrm{G}_A(\infty) + \mathrm{G}_B(\infty) = 1$.

2.5 Non-parametric Estimation of the Cumulative Risk Function

We assume no ties in the dataset (though the methods can be easily extended to account for them).

There are d event times $a_1 \dots a_d$, with $a_j < a_{j+1}$, and exactly one event, A or B, occurs at each event time. The marker ε_j indicates which event occurred at a_j ($\varepsilon_j \in \{A, B\}$) and there are r_j individuals in the risk set at that time. For notational convenience, $a_0 = 0$ by definition.

The Nelson-Aalen and Kaplan-Meier estimators for the joint event $J = A \cup B$ have the standard forms. The Nelson-Aalen estimate for the joint integrated hazard is given by:

$$\begin{aligned} \hat{\mathrm{H}}_J(t) &= \sum_{j:a_j \le t} \frac{1}{r_j} \\ &= \sum_{j:a_j \le t} \hat{h}_{J,j} \end{aligned}$$

where $\hat{h}_{J,j} = 1/r_j$. The Kaplan-Meier estimate for the joint survivor function is given in the same notation by:

$$\hat{\mathbb{P}}\left[t < T_J\right] = \hat{\mathrm{F}}_J(t) = \prod_{j:a_j \le t} \left(1 - \hat{h}_{J,j}\right) \tag{2}$$

and note that this is ultimately based on $\hat{\mathbb{P}}\left[a_{j-1} < T_J \le a_j | a_{j-1} < T_J\right] = \hat{h}_{J,j}$.

The estimate of the cause specific integrated hazard for A is simply the Nelson-Aalen estimator counting only events A:

$$\hat{\mathrm{H}}_A(t) = \sum_{j:a_j \le t} \frac{\mathbb{I}\left[\varepsilon_j = A\right]}{r_j} = \sum_{j:a_j \le t} \hat{h}_{A,j}$$

with

$$\hat{h}_{A,j} = \mathbb{I}\left[\varepsilon_j = A\right]/r_j = \hat{\mathbb{P}}\left[(a_{j-1} < T_J \le a_j)\&(T_J = T_A)|a_{j-1} < T_J\right]. \tag{3}$$

Note that $\hat{\mathrm{H}}_A(t) + \hat{\mathrm{H}}_B(t) = \hat{\mathrm{H}}_J(t)$.

An estimate for the cause specific risk function for event A is obtained by breaking down the risk function into the individual time intervals and summing the probability of event A occurring in that interval:

$$\hat{\mathrm{G}}_A(t) = \sum_{j:a_j \le t} \hat{\mathbb{P}}\left[(a_{j-1} < T_J \le a_j)\&(T_J = T_A)\right].$$

Each unconditional probability is the product of the probability that there is no event in $[0, a_{j-1}]$ multiplied by the probability of event A in $]a_{j-1}, a_j]$ conditional on no events in $[0, a_{j-1}]$:

$$\hat{\mathrm{G}}_A(t) = \sum_{j:a_j \le t} \hat{\mathbb{P}}\left[(a_{j-1} < T_J \le a_j)\&(T_J = T_A)|a_{j-1} < T_J\right] \hat{\mathbb{P}}\left[a_{j-1} < T_J\right]. \tag{4}$$

Note that we have just written down the non-parametric estimation equivalents of (1). The left hand factor of each term in (4) has been derived already (3) as $\hat{h}_{A,j}$ and the right hand factor is the Kaplan-Meier estimator (2) of $\mathrm{F}_J(t)$

evaluated at a_j:

$$\begin{aligned}\hat{\mathrm{G}}_A(t) &= \sum_{j:a_j \leq t} \hat{h}_{A,j} \prod_{j'=1}^{j-1} \left(1 - \hat{h}_{J,j'}\right). \\ &= \sum_{j:a_j \leq t} \frac{\mathbb{I}\left[\varepsilon_j = A\right]}{r_j} \prod_{j'=1}^{j-1} \left(1 - \frac{1}{r_{j'}}\right) \qquad (5)\end{aligned}$$

Expression (5) is known as the Aalen-Johansen estimate of the cumulative risk function.

Exercise Verify that $\hat{\mathrm{G}}_A(t) + \hat{\mathrm{G}}_B(t) + \hat{\mathrm{F}}_J(t) = 1$.