Competing Risks and Multistate Models

Henrique Laureano
http://leg.ufpr.br/~henrique

Last modification on 2020-03-09 20:29:20

Competing Risks and Multistate Models,
8nd chapter of *The Statistical Analysis of Failure Time Data*
Kalbfleisch and Prentice, 2002

Outline

- Competing Risks
  - Context & definitions
  - Likelihood functions
- Multistate Models

Notation & preliminaries,

- an underlying failure time T that may be subject to (independent) censoring
- a vector of possibly time-dependent covariates, $X(t) = \{x(u) : 0 \leq u < t\}$
- when failure occurs, it may be of any one of m distinct types or causes denoted by $J \in \{1, 2, \ldots, m\}$

As before, the overall rate or hazard function at time t is

$$\lambda[t; X(t)] = \lim_{h \to 0} h^{-1} \mathbb{P}[t \leq T < t + h \mid T \geq t, X(t)].$$

To model competing risks, we consider a type-specific or cause-specific hazard function or process

$$\lambda_j[t; X(t)] = \lim_{h \to 0} h^{-1} \mathbb{P}[t \leq T < t + h, J = j \mid T \geq t, X(t)].$$

for $j = 1, \ldots, m$ and $t > 0$.

type-specific hazard function, $\lambda_j[t; X(t)]$, in words,

$\lambda_j[t; X(t)]$ represents the instantaneous rate for failures of type j at time t given $X(t)$ and in the presence of all other failure types.

Without ties of the failure types,

$$\lambda[t; X(t)] = \sum_{j=1}^{m} \lambda_j[t; X(t)].$$

The overall survivor function is

$$F(t; X) = \mathbb{P}[T > t \mid X] = \exp\left\{ - \int_0^t \lambda(u; X) du \right\},$$

and the (sub)density function for the time to a type j failure is

$$\begin{aligned} f_j(t : X) &= \lim_{h \to 0} h^{-1} \mathbb{P}[t \leq T < t + h, J = j \mid X] \\ &= \lambda_j(t; X) F(t; X), \quad j = 1, \ldots, m. \end{aligned}$$

When the covariates are of the fixed or external type,

the likelihood on a sample subject to independent right censorship is written as a product of the survivor functions for the censored data and the subdensities for the observed failure times, i.e., the likelihood function can be written entirely in terms of the type-specific hazard functions.

The cumulative incidence function for type j failures corresponding to the external covariate is

$$\begin{aligned}\bar{F}_j(t; X) &= \mathbb{P}[T \leq t, J = j; X] \\ &= \int_0^t f_j(u; X) du, \quad t > 0,\end{aligned}$$

for $j = 1, \ldots, m$. Note that

$$p_j = \mathbb{P}[J = j] = \lim_{t \to \infty} \bar{F}_j(t; X), \quad j = 1, \ldots, m$$

and $\sum_{j=1}^{m} p_j = 1$.

$F_j(t; X)$ has no simple probability interpretation within the competing risks model, at least not without introducing strong additional assumptions.

Example 8.1. Suppose that $m = 2$ and that the covariate is a treatment indicator $x = 0, 1$.

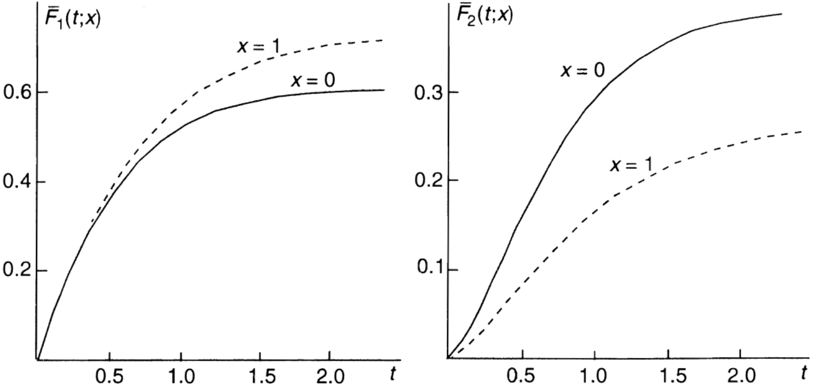

Figure 8.1 Cumulative incidence functions for Example 8.1.

Likelihoods

Consider $\{t_i, \delta_i, j_i, X_i(t_i)\}_{i=1}^{n}$.

If the censoring is independent, the likelihood (or partial likelihood) is proportional to

$$
\begin{aligned}
L &= \prod_{i=1}^{n} \left(\{\lambda_{j_i}[t_i; X_i(t_i)]\}^{\delta_i} \prod_{j=1}^{m} \exp\left\{ -\int_0^{t_i} \lambda_j[u; X_i(u)]du \right\} \right) \\
&= \prod_{j=1}^{m} \left(\{\lambda_j[t_i; X_i(t_i)]\}^{\delta_{ji}} \exp\left\{ -\int_0^{\infty} \sum_{i=1}^{n} Y_i(t)\lambda_j[t; X_i(t)]dt \right\} \right).
\end{aligned}
$$

Any of the methods of preceding chapters can be used for inference about the $\lambda_j[t; X(t)]$'s.

Likelihoods

We can also generalize simple explanatory methods such as Kaplan-Meier and Nelson-Aalen estimators to competing risks data.

Let $t_1 < t_2 < \cdots < t_k$ denote the k distinct failure times for all failure types combined. Then, the likelihood function can be written

$$L = \prod_{i=1}^{k} \left(\prod_{j=1}^{m} \{[F_j(t_i^-) - F_j(t_i)]F(t_i^-)\}^{d_{ji}} \prod_{l=1}^{C_i} [F(t_{il})]^{c_{il}} \right).$$

Its nonparametric MLE places mass only at the observed failure times $1, \ldots, k$, so the partially maximized likelihood can be rewritten using expressions for discrete models, to obtain

$$\text{Multinomial likelihood :} \quad \hat{L} = \prod_{i=1}^{k} \left[\prod_{j=1}^{m} \lambda_{ji}^{d_{ji}} (1 - \lambda_i)^{n_i - d_i} \right].$$

Maximizing it gives the MLE $\hat{\lambda}_{ji} = d_{ji}/n_i$.

Likelihoods

The cumulative hazard function is then estimated by
$\hat{\Lambda}_j(t) = \sum_{i=1}^{k} \mathbf{1}(t_i \leq t) d_{ji}/n_i$, $t \geq 0$.

- This yields the Nelson-Aalen estimate of the total cumulative hazard and the Kaplan-Meier estimate of the overall survivor function $F(t)$.

The estimated cumulative incidence function is also discrete, and is given by

$$\hat{\hat{F}}_j(t) = \sum_{\{i | t_i \leq t\}} d_{ji} n_i^{-1} \hat{F}(t_i^-), \quad j = 1, \dots, m.$$

Likelihoods

Consider now a relative risk or Cox model for the
cause-specific hazard functions

$$\lambda_j[t; X(t)] = \lambda_{0j}(t) \exp\{Z(t)^\top \beta_j\}, \quad j = 1, \ldots, m.$$

The corresponding partial likelihood is

$$L(\beta) = \prod_{j=1}^{m} \prod_{i=1}^{k_j} \frac{\exp\{Z_{ji}(t_{ji})^\top \beta_j\}}{\sum_{l \in R(t_{ji})} \exp\{Z_l(t_{ji})^\top \beta_j\}}.$$

If applicable, a proportional risks model

$$\lambda_j[t; X(t)] = \lambda_0(t) \exp\{\gamma_j + Z(t)^\top \beta_j\}, \quad j = 1, \ldots, m,$$

would yield more efficient β_j estimators, in which the
cause-specific hazards are assumed to be proportional
to each other (for uniqueness set $\gamma_1 = 0$).

Likelihoods

The partial likelihood of the proportional risk model can then be written

$$\prod_{i=1}^{k} \frac{\exp\{\gamma_{j_i} + Z_i(t_i)^\top \beta_{j_i}\}}{\sum_{j=1}^{m} \sum_{l=1}^{n} Y_l(t_i) \exp\{\gamma_{j+} Z_l(t_i)^\top \beta_j\}}.$$

As is the general relative risk model,
an adjustment is needed to handle tied failure times.

Although it would often be more restrictive than is desirable, the proportial risk model has some attractive properties. For instance, the probability that an individual with fixed covariate Z has failure type j is

$$\mathbb{P}[J = j; Z] = \frac{\exp\{\gamma_j + Z^\top \beta_j\}}{\sum_{h=1}^{m} \exp\{\gamma_h + Z^\top \beta_h\}}, \quad j = 1, \ldots, m,$$

regardless of $\lambda_0(\cdot)$.

Example 8.2.

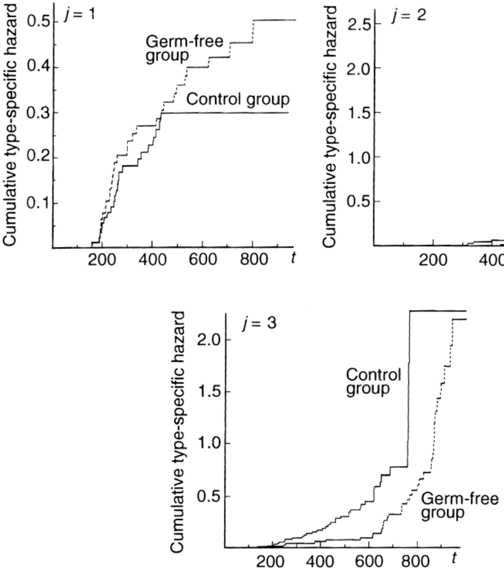

Figure 8.2 Estimates of the cumulative type-specific hazard functions for the data of Example 8.2.

Example 8.2.

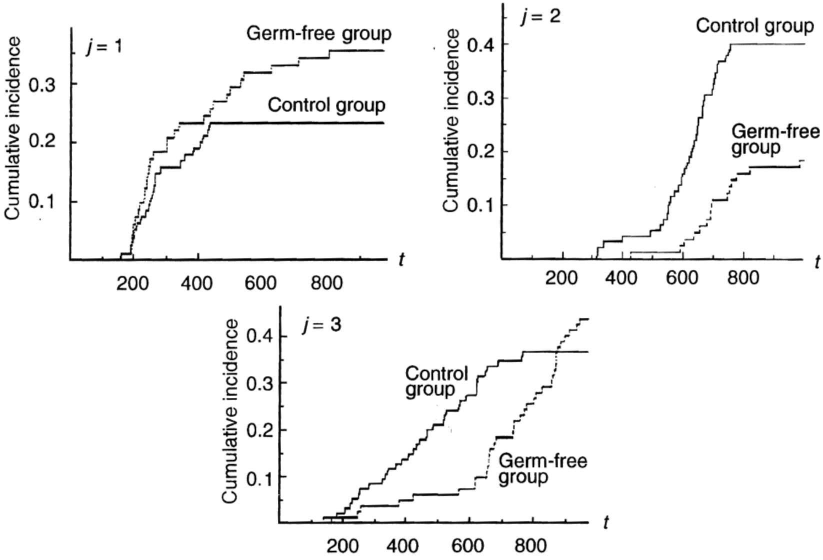

Figure 8.3 Estimates of the cumulative incidence functions (8.11) for the data of Example 8.2.

Rebolledo's Theorem

Idea: Consider conditions under which the distribution of $U^{(n)}$ approaches a normal limit as $n \to \infty$.

[Rebolledo's] Let t be a fixed time in $[0, \tau]$ and consider the conditions:

(a) $\langle U^{(n)} \rangle(t) \xrightarrow{\mathcal{P}} V(t)$ as $n \to \infty$.

(b) $[U^{(n)}](t) \xrightarrow{\mathcal{P}} V(t)$ as $n \to \infty$

(c) $\langle U_{\epsilon j}^{(n)} \rangle(t) \xrightarrow{\mathcal{P}} 0$ as $n \to \infty$, $\forall j = 1, \ldots, k$ and $\epsilon > 0$.

Then either (a) or (b) together with (c) imply that $U^{(n)}(t) \xrightarrow{\mathcal{D}} N(0, V(t))$.

Let $V(t)$ be a $k \times k$ positive semidefinite matrix on the interval $[0, \tau]$ where $V(0) = 0$ and $V(t) - V(s)$ is positive semidefinite $\forall s, t$ satisfying $0 \leq s \leq t \leq \tau$.

General version of Rebolledo's Theorem

General? Results that involve the joint distribution of $[U^{(n)}(t_1), \ldots, U^{(n)}(t_n)]$ or the convergence of the entire process over the specified interval.

[theorem] Extend conditions (a), (b), and (c) so that the convergence holds uniformly $\forall t \in K$ where $K \subseteq (0, \tau]$. If the resulting (c) together with either (a) or (b) holds, then

$$[U^{(n)}(t_1)^\top, \ldots, U^{(n)}(t_r)^\top]^\top \xrightarrow{\mathcal{D}} N(0, \Sigma),$$

where Σ is a $(kr) \times (kr)$ matrix comprised of $k \times k$ blocks. Thus

$$\Sigma = \begin{bmatrix} V(t_1) & V(t_1) & \cdots & V(t_1) \\ V(t_1) & V(t_2) & \cdots & V(t_2) \\ \vdots & \vdots & & \vdots \\ V(t_1) & V(t_2) & \cdots & V(t_r) \end{bmatrix}.$$

Further, if $K = (0, \tau]$, then $U^{(n)}$ converges weakly on K to a k-variate Gaussian martingale with covariance function $V(t)$.

Asymptotics: Cox model

Under independent right censoring, the Cox model gives

$$\mathbb{P}[dN_i(t) = 1 | \mathcal{F}_{t^-}] = Y_i(t)\lambda_i(t) = Y_i(t)\exp\{Z_i(t)^\top\beta\}\lambda_0(t)dt.$$

Note that $\lambda_i(t)$ is the intensity or hazard function for the underlying uncensored counting process $\tilde{N}_i(t)$.

Under some boring and too mathly conditions, the partial likelihood estimatior $\hat{\beta}$ obtained by maximizing the log partial likelihood

$$l(\beta,\tau) = \sum_{i=1}^{n}\int_0^\tau Z_i(u)^\top\beta dN_i(u) - \int_0^\tau \log\left[\sum_{j=1}^{n} Y_j(u)\exp\{Z_j(u)^\top\beta\}\right] dN_\cdot(u)$$

is consistent for β.

Asymptotics: Cox model

In the case of no ties, the score function, U, based on the partial likelihood, is a mean 0 vector-valued martingale wrt $\mathcal{F}_t$ and can be written as

$$U(\beta, t) = \sum_{i=1}^{n} \int_0^t \left[Z_i(u) - \sum_{l=1}^{n} Z_l(u) \frac{Y_l(u) \exp\{Z_l(u)^\top \beta\}}{\sum_{j=1}^{n} Y_j(u) \exp\{Z_j(u)^\top \beta\}} \right] dM_i(u).$$

Under some more boring and too mathly conditions, as $n \to \infty$,

$$n^{-1/2} U(\beta, \tau) \xrightarrow{\mathcal{D}} N(0, \Sigma(\beta, \tau)),$$

$$n^{1/2}(\hat{\beta} - \beta) \xrightarrow{\mathcal{D}} N(0, \Sigma(\beta, \tau)^{-1}), \text{ with}$$

$$\Sigma(\beta, \tau) = \int_0^\tau \left[\sum_{i=1}^{n} Y_i(u) Z_i(u) Z_i(u)^\top \exp\{Z_i(u)^\top \beta\} - \frac{\left[\sum_{i=1}^{n} Y_i(u) Z_i(u) \exp\{Z_i(u)^\top \beta\}\right]^{\otimes 2}}{\sum_{i=1}^{n} Y_i(u) \exp\{Z_i(u)^\top \beta\}} \right] \lambda_0(u) du,$$

$t \in [0, \tau]$.

Asymptotics: parametric models

✓ Asymptotic results apply to independent right censoring and left truncation.

✗ They do not cover interval censoring or right truncation.

And again, we have

$$\mathbb{P}[dN_i(t) = 1|\mathcal{F}_{t^-}] = Y_i(t)\lambda_i(t|\theta)dt, \quad i = 1, \dots, n.$$

The log-likelihood function arising from continuous failure time data on the interval $(0, \tau]$ can be written

$$l(\theta) = \sum_{i=1}^{n} \int_0^\tau \log \lambda_i(t|\theta) dN_i(t) - \sum_{i=1}^{n} \int_0^\tau Y_i(t)\lambda_i(t|\theta)dt.$$

Asymptotics: parametric models

The score process on data on $(0, \tau]$ is

$$U(\theta, t) = \sum_{i=1}^{n} \int_0^t \left[\frac{\partial}{\partial \theta} \log \lambda_i(u|\theta) \right] dM_i(u), \quad 0 < t < \tau$$

where $M_i(t)$ are orthogonal martingales.

Via central limit for the score function
(and some other conditions, of course), as $n \to \infty$,

$$n^{-1/2} U(\theta, \tau) \xrightarrow{\mathcal{D}} N(0, \Sigma(\theta)),$$

$n^{1/2}(\hat{\theta} - \theta) \xrightarrow{\mathcal{D}} N(0, \Sigma(\theta)^{-1})$, with $\Sigma(\theta)$ est. consistently by $n^{-1} I(\hat{\theta})$.

$$I(\theta) = \sum_{i=1}^{n} \int_0^{\tau} \left[\frac{\partial}{\partial \theta} \log \lambda_i(u|\theta) \right]^{\otimes 2} dN_i(u) - \sum_{i=1}^{n} \int_0^{\tau} \frac{\partial^2 \lambda_i(u|\theta)}{\partial \theta \partial \theta^{\top}} \lambda_i(u|\theta)^{-1} dM_i(u).$$

Efficiency of the Cox model estimator, $\hat{\beta}_{\text{cox}}$

» With a hazard function specified up to a certain unknown parameter α, e.g., $\lambda_0(t) = \alpha h_0(t)$, which approach would lead to more precise asymptotic estimation of β, $\hat{\beta}_{\text{cox}}$ or a MLE $\hat{\beta}$?

If Z is time independent, $\hat{\beta}_{\text{cox}}$ is then asymptotically fully efficient;
If Z is time dependent, it is not.

Why?

The average Z value over the risk set varies with time in a way that such variations introduce asymptotic correlations between the estimator of β and that of α.


THANK YOU
memegenerator.net