Competing Risks and Multistate Models

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

Last modification on 2020-03-09 10:28:43

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

Outline

» Competing Risks

A counting process $N = \{N(t), t \geq 0\}$

is a stochastic process with $N(0) = 0$ and whose value at time t counts the number of events that have occured in the interval $(0, t]$.

- The sample paths of N are nondecrising step functions that jump whenever an event (or events) occur.
- In continuous time,

  no two counting processes can jump at the same time.

- In discrete time, they can.

Number of events that occur in the interval $[t, t + dt)$?
$dN(t) = N(t^- + dt) - N(t^-)$.

Number of events that occur at time t? $\Delta N(t) = N(t) - N(t^-)$.

And what about more general counting processes where individuals may experience more than one event? Chapters 8, 9, and 10.

Filtration: history of events

observed counting process: $N_i = \{N_i(t), t \geq 0\}$

underlying counting process: $\tilde{N}_i = \{\tilde{N}_i(t), 0 \leq t\}$, $\quad \tilde{N}_i(t) = \mathbf{1}(T_i \leq t)$

at-risk process: $\{Y_i(t), t \geq 0\}$, $\quad Y_i(t) = \mathbf{1}(T_i \geq t, C_i \geq t)$

key concept: filtration

$$\mathcal{F}_t = \sigma\{N_i(u), Y_i(u^+), X_i(u^+), i = 1, \ldots, n; 0 \leq u \leq t\}, \quad t > 0,$$

where

$Y_i(u^+) = \lim_{s \to u^+} Y_i(s)$;

stochastic time-dependent covariate: $X_i(t) = \{x_i(u) : 0 \leq u \leq t\}$.

The notation $\sigma[\cdot]$ specifies the sigma algebra of events generated by the variables given in the brackets.

Intensity functions

The intensities or rates for the processes N_i are defined with reference to the filtration $\mathcal{F}_t$. If the censoring process is independent, the intensity model for the counting process N_i is

$$\mathbb{P}[dN_i(t) = 1 | \mathcal{F}_{t^-}] = Y_i(t) d\Lambda_i(t), \quad i = 1, \dots, n, \quad t > 0.$$

The hazard model can be written $d\Lambda_i(t) = \mathbb{P}[d\tilde{N}_i(t) = 1 | X_i(t), \tilde{N}_i(t^-) = 0]$.

Λ_i is called the cumulative intensity process of the counting process $\tilde{N}_i$.

- In the continuous case, $\mathbb{P}[dN_i(t) = 1 | \mathcal{F}_{t^-}] = Y_i(t) \lambda_i(t) dt$
- In the discrete case, $\mathbb{P}[dN_i(a_l) = 1 | \mathcal{F}_{a_l^-}] = Y_i(a_l) \lambda_{il}, \quad l = 1, 2, \dots$

$\lambda_i(t)$ and λ_{il} are the corresponding intensity processes.

Martingales: Intro

$$
\begin{aligned}
M_i(t) &= N_i(t) - \int_0^t Y_i(u)\lambda_i(u)du, \quad t \geq 0. \\
&= \int_0^t dM_i(u), \\
dM_i(t) &= dN_i(t) - Y_i(t)\lambda_i(t)dt.
\end{aligned}
$$

If

- $\mathbb{E}[dM_i(t)|\mathcal{F}_{t-}] = 0, \quad \forall t; \quad \equiv \quad \mathbb{E}[M_i(t)|\mathcal{F}_s] = M_i(s), \quad \forall s \leq t.$

Then, $M_i(t)$ is a martingale.

Consequences:

- $\mathbb{E}[M_i(t)] = 0, \quad \forall t;$
- the process $M_i(t)$ has uncorrelated increments, i.e., $\mathbb{E}[(M_i(t) - M_i(s)) \times M_i(s)] = 0, \quad \forall 0 < s < t.$

Decomposing $N_i(t)$ into two processes

$$N_i(t) = \underbrace{\int_0^t Y_i(u)\lambda_i(u)du}_{\substack{\text{compensator of the} \\ \text{counting process } N_i \\ \text{wrt the filtration } \mathcal{F}_t}} + \underbrace{M_i(t)}_{\substack{\text{counting} \\ \text{process} \\ \text{martingale} \\ \text{corresponding} \\ \text{to } N_i(t)}}$$

$$dN_i(t) = Y_i(t)\lambda_i(t)dt + dM_i(t).$$

In the discrete case, the discrete-time martingale is

$$\begin{aligned} N_i(t) &= \int Y_i(u)d\Lambda_i(u) + M_i(t) \\ &= \sum_{a_l \leq t} Y_i(a_l)\lambda_{il} + M_i(t), \end{aligned}$$

$$dN_i(a_l) = Y_i(a_l)\lambda_{il} + dM_i(a_l).$$

More about martingales

In essense, a martingale is a process that has no drift and whose increments are uncorrelated.

- We say that $M(t)$ is a mean zero martingale if $\mathbb{E}[M(0)] = 0$, and hence $\mathbb{E}[M(t)] = 0, \forall t$.
- The martingale $M(t)$ is said to be square integrable (or have finite variance) if $\mathbb{E}[M^2(t)] = \mathbb{V}[M(t)] < \infty, \forall t \leq \tau$.

It is useful to define two technical terms applied to a stochastic process $U = \{U(t), t \geq 0\}$.

Adapted

U is said to be adapted to the filtration $\mathcal{F}_t$ if $U(t)$ is $\mathcal{F}_t$ measurable for each $t \in [0, \tau]$, i.e., the value of $U(t)$ is fixed once $\mathcal{F}_t$ is given.

Predictable

U is said to be predictable wrt the filtration $\mathcal{F}_t$ if $U(t)$ is $\mathcal{F}_{t^-}$ measurable for all $t \in [0, \tau]$, i.e., the value of $U(t)$ is fixed once $\mathcal{F}_{t^-}$ is given.

More about martingales

The process $\{\bar{M}(t), 0 \leq t \leq \tau\}$ is a submartingale wrt $\mathcal{F}_t$ if it is adapted and satisfies

$$\mathbb{E}[\bar{M}(t)|\mathcal{F}_s] \geq \bar{M}(s) \quad \forall s \leq t \leq \tau.$$

- A counting process $N(t)$ is a submartingale.

Predictable variation process

The predictable variation process of a square-integrable martingale M is

$$\langle M\rangle(t) = \int_0^t \mathbb{V}[dM(u)|\mathcal{F}_{u^-}].$$

Equivalently, $d\langle M\rangle(t) = \mathbb{V}[dM(u)|\mathcal{F}_{u^-}]$.

In statistical terms, the primary role of the predictable variation process is that for given t, $\langle M\rangle(t)$ provides a systematic approach to estimating the variance of $M(t)$.

Variance of $M(t)$

$$\mathbb{V}[M(t)] = \mathbb{E}[M^2(t)] = \mathbb{E}[\langle M\rangle(t)]$$

and $\langle M\rangle(t)$ is an unbiased estimator of $\mathbb{V}[M(t)]$.

Usually, $\langle M\rangle(t)$ involves the parameters of the model.

There is an alternative estimator of $\mathbb{V}[M(t)]$ that in some problems is a function of observed quantities only. This is the quadratic variation or optional variation process $[M](t)$.

$$[M](t) = \sum_{s\leq t}(\Delta M(s))^2.$$

$[M](t)$ also provides an unbiased estimator of $\mathbb{V}[M(t)]$.

Continuous-time counting processes

For each i, the process

$$M_i(t) = N_i(t) - \int_0^t Y_i(u)\lambda_i(u)du$$

is a mean zero martingale wrt the filtration $\mathcal{F}_t$. The corresponding predictable variation process is

$$\langle M \rangle(t) = \int_0^t Y_i(u)\lambda_i(u)du$$

since $d\langle M \rangle(u) = \mathbb{V}[dM(u)|\mathcal{F}_{u^-}] = \mathbb{E}[dN_i(u)|\mathcal{F}_{u^-}] = Y_i(u)\lambda_i(u)du$, $0 \leq u \leq \tau$. The optional variation process is

$$[M_i](t) = N_i(t).$$

Both $\langle M_i \rangle(t)$ and $[M_i](t)$ provide unbiased estimates of $\mathbb{V}[M_i(t)]$, but only the latter is a function of the data only.

Discrete-time continuous processes

In discrete time we have

$$\mathbb{P}[dN_i(t) = 1 | \mathcal{F}_{t^-}] = \mathbb{E}[dN_i(t) | \mathcal{F}_{t^-}] = Y_i(t) d\Lambda_i(t),$$

where $\Lambda_i(t) = \sum_{a_l \leq t} \lambda_{il}$ and λ_{il} are the discrete hazard probabilities.

The corresponding martingale is

$$M_i(t) = N_i(t) - \int_0^t Y_i(u) d\Lambda_i(u)$$

and the predictable variation process is

$$\langle M_i \rangle (t) = \sum_{a_l \leq t} Y_i(a_l)(1 - \lambda_{il})\lambda_{il}.$$

The increment $d\langle M_i \rangle(t)$ at a_l corresponds to
$\mathbb{V}[dN_i(a_l) | \mathcal{F}_{t^-}] = Y_i(a_l)(1 - \lambda_{il})\lambda_{il}$.

Martingale central limit theorem: Rebolledo (1980)

The central limit theorem due to Rebolledo (1980) is the main result useful for asymptotics in the applications of the book. This theorem drew together earlier work on central limit theorems for discrete martingales and gave a general version applicable to martingales arising from counting processes in discrete time, continuous time, or a mixture of the two.

What do we need? A vector of k martingales, $U^{(n)}$, where

$$U_j^{(n)}(t) = \sum_{i=1}^{r_n} \int_0^t G_{ji}^{(n)}(s) dM_i^{(n)}(s), \quad j = 1, \ldots, k.$$

$G_{ji}^{(n)}(t)$ are predictable processes wrt a filtration $\mathcal{F}_t^{(n)}$.

For given $\epsilon > 0$, we define a vector of related processes,

$$U_{\epsilon j}^{(n)}(t) = \sum_{i=1}^{r_n} \int_0^t G_{ji}^{(n)}(u) \mathbf{1}(|G_{ji}^{(n)}(u)| > \epsilon) dM_i^{(n)}(u),$$

$j = 1, \ldots, k$, $t \in [0, \tau]$, which registers all jumps of size ϵ or more in the original component processes $U_j^{(n)}$.

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) $\langle U^{(n)}_{\epsilon j} \rangle(t) \xrightarrow{\mathcal{P}} 0$ 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 $[U^{(n)}(t_1), \dots, U^{(n)}(t_n)]$ 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

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

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