Time-dependent covariates

Rasmus Waagepetersen

November 18, 2024

Martingale approach to Cox proportional hazards

We can write Cox partial likelihood with time-varying covariate as

$$L(\beta) = \prod_{i \in D} \frac{\exp[\beta^\mathsf{T} Z_i(t_i)]}{\sum_{l=1}^{n} Y_l(t_i) \exp[\beta^\mathsf{T} Z_l(t_i)]}$$

where Y_l is 'at risk' process for lth individual and Z_l is covariate *process* for lth individual.

Score process for data *up to time* t:

$$u(\beta, t) = \sum_{i \in D : t_i \le t} (Z_i(t_i) - E(t_i))$$

We verified last time that score process is a martingale ($\Rightarrow$ asymptotic normality for $u(\beta, t)/\sqrt{n}$) and that variance of score process is equal to Fisher information. This is background for result

$$\hat{\beta} \approx N(\beta, i(\beta)^{-1})$$

Time-dependent covariates

Our excursion into the realm of counting process and martingales showed that it poses no problems to introduce predictable random time-varying covariates in the Cox model.

Reasons for doing so: the value of a covariate at time $t = 0$ may not be relevant - instead the hazard at a given time t depends on the *current* value of the covariate at time t.

Example: cumulative power produced for a windturbine as a function of time $\mathrm{gwh}(\cdot)$ may be a proxy for wear of the windturbine. Hence the hazard should depend at each time t on $\mathrm{gwh}(t)$.

Why wrong to use $\mathrm{gwh}(t_i)$ as fixed covariate ?

Example from Therneau (survival in relation to cumulated dose of medication): use of dose at time of death is wrong - to get a big dose you have to live long. If hazard is completely unrelated to dose we would still see high dose associated with long survival.

Internal vs. external covariates

Some covariates are external in the sense that they exist/develop independently of the survival of a patient.

Example: air pollution and survival to death of respiratory disease.

Other covariates only 'exist'/can be recorded as long as the patient is alive - e.g. blood pressure measured over time. These are called internal covariates.

For fitting of a Cox regression model the distinction between external and internal covariates is not important.

However, the distinction matters when it comes to predicting survival - next slide.

Prediction

Suppose we are able to predict the value of a covariate $Z(t)$ for any $t \geq 0$. Then we can define the distribution of the survival time conditional on $Z = \{Z(t)\}_{t \geq 0}$ by the conditional survival function

$$S(t|Z) = \exp(-\int_0^t h_0(u) \exp(\beta^\mathsf{T} Z(u)) \mathrm{d}u)$$

This may in principle be possible for external covariates if we can solve the prediction problem (which is not straightforward).

The situation is more complicated for internal covariates. Here a hierarchical specification may not make sense since e.g. blood pressure can only be measured as long as the patient is alive - which depends on the lifetime X which again depends on $Z(t)$, $0 \leq t \leq X$.

One approach for internal variables could be to adopt process point of view and simulate simultaneously $N(t)$ and $Z(t)$ ahead in time until $N(t) = 1$.

Cox partial likelihood

Cox proportional likelihood compares risk for the group of patients at risk at a specific death time. We should thus use the values of the covariates that are appropiate for each patient at risk at that specific time. E.g. not future values of a time-dependent covariate whose value depend on duration of survival.

What about blood pressure measured at time $t = 0$?

Valid since for patients being compared at time t it is the same covariate (bloodpressure measured t time units ago) - but blood pressure at time t may be a better predictor of hazard at time t.

Example from KM: disease-free survival improves after platelet (blodplader) recovery. This recovery happens at a random time after time of transplation. Should we just use indicator for whether recovery was observed as covariate ?

Test for proportional hazards

Given covariate z fit model with z and time-dependent version of z, $z(t) = z\log(t)$. Then hazard is

$$h_0(t)\exp(\beta_1 z + \beta_2 z \log t) = h_0(t)\exp(\beta_1 z)t^{\beta_2 z}$$

and hazard ratio for subjects with covariate values z_1 and z_2 is

$$\exp(\beta_1(z_2 - z_1))t^{\beta_2(z_2 - z_1)}$$

That is, hazard ratio can be increasing or decreasing as a function of time depending on sign of $\beta_2(z_2 - z_1)$.

NB since z is given and fixed for a patient, it is more appropriate to talk about a *time-varying effect* of z:

$$\beta_1 z + \beta_2 z \log t = (\beta_1 + \beta_2 \log t)z = \beta(t)z$$

where $\beta(t) = \beta_1 + \beta_2 \log t$

Age as a time-dependent variable ?

Exercise: show that using the timedependent covariate $z_i(t) = a_i + t$ for the ith subject in a Cox regression is the same as using age a_i at $t = 0$ as a fixed covariate.

Implementation

In R two options (see vignette by Therneau et al.):

- specify intervals where time-dependent variable takes a certain value.
- use `tt` functionality.

Example from KM Section 9.2 - implementation in R

See R-code.