Survival Analysis with Functional Covariates

T Padma Ragaleena

25 April 2023

Survival Analysis

- Outcome variable of interest = **time until an event occurs**
- Examples: time until patient dies after surgery, time until patient recovers from illness etc.
- Outcome variable is referred to as **survival time**
- I will continue to use the example of time until death.

What is different about time to event data?

1. Interested in: whether the event occurred AND when the event occurred
2. Traditional methods not equipped to handle **censoring** - a special kind of missingness observed in time-to-event data analysis.
3. Other reasons for censoring: lost to follow up, withdraws from study

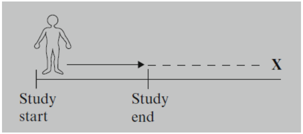

Figure: Censored patient

Cox Proportional Hazards (PH) model

- Cox PH model given real valued predictors $\mathbf{X} = (X_1, \cdots, X_p)$

$$h(t, \mathbf{X}) = h_0(t) \exp\{\sum_{i=1}^{p} \beta_i X_i\}$$

- $h(t, \mathbf{X}) =$ hazard at time t for individual with predictors $\mathbf{X}$
- $h_0(t)$ represents baseline hazard - unspecified function.
- Assumes the **Proportional hazards assumption**
- **Semi-parametric model** - one of the reasons why the model is so popular. Can estimate all β_i and the hazard function without knowing the form of $h_0(t)$.
- Observe: no distributional assumption for the outcome variable.

Cox likelihood/ Partial likelihood

- We want to ensure we include the *partial* information until censorship as well.
- Here, we do not model the joint distribution of the outcomes.
- Likelihood depends on **order of events** and independent of baseline hazard.

ID	TIME	STATUS	SMOKE
Barry	2	1	1
Gary	3	1	0
Harry	5	0	0
Larry	8	1	1

SURVT = Survival time (in years)
STATUS = 1 for event, 0 for censorship
SMOKE = 1 for a smoker, 0 for a nonsmoker

Cox PH model

$$h(t) = h_0(t)e^{\beta_1 SMOKE}$$

Likelihood is product of 3 terms

$$L = L_1 \times L_2 \times L_3$$

$$L_1 = \left[\frac{h_0(t)e^{\beta_1}}{h_0(t)e^{\beta_1} + h_0(t)e^0 + h_0(t)e^0 + h_0(t)e^{\beta_1}}\right]$$

$$L_2 = \left[\frac{h_0(t)e^0}{h_0(t)e^0 + h_0(t)e^0 + h_0(t)e^{\beta_1}}\right]$$

$$L_3 = \left[\frac{h_0(t)e^{\beta_1}}{h_0(t)e^{\beta_1}}\right]$$

Incorporating functional covariates

- First paper to incorporate a functional predictor to model time-to-even data is *Generalized linear models with functional predictors* (2002) by G.M.James

$$E[Y] = g(\mu) = \_0 + \gamma z + \int w_1(t)X(t)\mathrm{d}t$$

- Assume each predictor can be modelled as a smooth curve using cubic splines.
- Similar idea has been used to incorporate functional predictors into the Cox-PH model.

$$\log[h_i(t, Z_i; \gamma, \beta(.))] = \log h_0(t) + \gamma^T Z_i + \int X_i(s)\beta(s)\mathrm{d}s$$

- I will give an over-view of two approaches that I came across:
  1. First approach is similar to Ridge regression.
  2. The second approach is similar to dimension reduction using PCA.

Approach 1: Penalized partial likelihood

- Approximating $\beta(s)$ using B-spline basis $\{\phi_i(s)\}_{i=1}^{k_b}$ gives

$$\beta(s) = \sum_{i=1}^{k_b} b_i \phi_i(s)$$

- Using this approximation in the functional Cox PH model gives:

$$\log[h_i(t, Z_i; \gamma, \beta(.))] = \log h_0(t) + \gamma^T Z_i + \mathbf{b}^T \mathbf{c}_i \text{ where } c_{ij} = \int X_i(s)\phi_j(s)\mathrm{d}s$$

- Penalized partial likelihood with penalty function $p(\theta)$ for $\theta^T = (\gamma^T b^T)$, $\lambda \geq 0$ and symmetric, positive semi-definite matrix.

$$l^\lambda(\theta) = l(\theta) - \frac{1}{2}\lambda p(\theta) \text{ where } p(\theta) = \theta^T D\theta$$

- Given λ, estimate θ by maximizing partial log-likelihood using **Newton-Raphson procedure**.

Approach 1: Penalized partial likelihood

- **Cross-validated log likelihood (CVL)** is used to find the “optimal" smoothing parameter.
- Define the following:

$$\widehat{\theta}^{\lambda}_{(-i)} = \mathrm{argmax}_{\theta} l^{(p)}_{\lambda,(-i)}(\theta) \text{ and } l^{(p)}_{\lambda,i}(\cdot) = l^{(p)}_{\lambda}(\cdot) - l^{(p)}_{\lambda,(-i)}(\cdot)$$

- Then we define:

$$CVL(\lambda) = \sum_{i=1}^{n} l^{(p)}_{\lambda,i}(\widehat{\theta}^{\lambda}_{(-i)})$$

- Since the above equation is computationally expensive, authors define a different AIC based approximation of CVL that can be used to compute λ.

Approach 2: Functional PCA

- The model is still the same:

$$\log[h_i(t, Z_i; \gamma, \beta(.))] = \log h_0(t) + \gamma^T Z_i + \int X_i(s)\beta(s)\mathrm{d}s$$

- Here, we model $\{X(t)\}_{t\in\mathcal{T}}$ as a stochastic process with mean $\mu(t)$ and covariance matrix $G(s,t) = cov(X(s), X(t))$.
- Think of the operator $\widehat{A}$ as a map between function spaces. Eg. Differential operator

$\widehat{A}f = k \cdot f$ then f is the **eigen function** and k is **eigen value**

- Let $\lambda_1 \geq \lambda_2 \geq \cdots$ and $\phi_1, \phi_2, \cdots$ be the eigen values and eigen functions of covariance function.

Approach 2: Functional PCA

- Assume the following result: Centered Stochastic process can be spanned by the eigen function basis.

$$X(t) - \mu(t) = \sum_{k=1}^{\infty} \xi_k \phi_k(t) \text{ where } \xi_k \text{ is k-th PC associated with } \phi_k$$

- Common approximation used (just like in PCA):

$$X(t) - \mu(t) \approx \sum_{k=1}^{r_n} \xi_k \phi_k(t)$$

- The above approximation has theoretical convergence gurantee.
- Authors describe how to choose “m".

Approach 2: Functional PCA

- Functional Cox PH model:

$$\log[h_i(t, Z_i; \gamma, \beta(.))] = \log h_0(t) + \gamma^T Z_i + \int X_i(s)\beta(s)\mathrm{d}s$$

- Substitute following two expressions in the main equation:

$$\beta(s) = \sum_{i=1}^{k_b} b_i \phi_i(s) \text{ and } X(t) = \mu(t) + \sum_{k=1}^{m} \xi_k \phi_k(t)$$

- Using the fact that the eigen basis form an orthonormal set, we can simplify the main equation as follows:

$$\log[h_i(t, Z_i; \gamma, \beta(.))] = \log h_0(t) + \gamma^T Z_i + \sum_{j=1}^{r_n} \xi_{ij}\beta_j$$

Approach 2: Functional PCA

- Functional Cox PH model

$$\log[h_i(t, Z_i; \gamma, \beta(.))] = \log h_0(t) + \gamma^T Z_i + \sum_{j=1}^{r_n} \xi_{ij}\beta_j$$

- Authors then describe how to estimate $h_0(t)$, (γ^T, β^T) after getting smoothed estimates for $X_i(s)$ for each i.

References

1. Gellar JE, Colantuoni E, Needham DM, Crainiceanu CM. **Cox regression models with functional covariates for survival data.** **Statistical Modelling**. 2015;15(3):256-278.
2. Kong, D., Ibrahim, J. G., Lee, E. and Zhu, H. (2017). **FLCRM: Functional linear cox regression model.** **Biometrics** 74, 109–117.
3. James, G. M. (2002). **Generalized Linear Models with Functional Predictors**. **Journal of the Royal Statistical Society Series B: Statistical Methodology** 64, 411–432.
4. Eilers, P. H. C. and Marx, B. D. (1996). **Flexible smoothing with B-splines and penalties**. **Statistical Science** 11. doi:10.1214/ss/1038425655.
5. Verweij, P. J. M. and Van Houwelingen, H. C. (1994). **Penalized likelihood in Cox regression**. **Statistics in Medicine** 13, 2427–2436.
6. Balan, T. A. and Putter, H. (2020). **A tutorial on frailty models**. **Statistical Methods in Medical Research** 29, 3424–3454.
7. Kleinbaum, D. G. and Klein, M. (2012). **Survival Analysis**. Springer New York. doi:10.1007/978-1-4419-6646-9.