Exercises on clustered survival data

Advanced survival analysis 2018 day 5

Exercise 1 : Diabetic retinopathy study

Consider the data set `diabetes` from the `timereg` package. You can load it by

```
library(timereg)
data(diabetes)
```

The dataset contains 197 patients with "high-risk" diabetic retinopathy, which is a leading cause of blindness in patients under 60 years of age. Each patient had one eye randomized to laser treatment and the other eye received no treatment. For each eye, the event of interest was the time from treatment to blindness. Each patient is represented by two lines:

- `time:` time from treatment to blindness
- `status:` 1 blind, 0 censored
- `treat:` laser treatment
- `adult:` type of diabetes (adult/juvenile)
- `id:` person id

Are blindness times correlated within persons? If so, how dependent? Does laser treatment affect blindness time? Does type of diabetes affect blindness onset? In all steps below report the hazard ratio and confidence interval associated with treatment and type of diabetes. Which model do you prefer? How do they differ in interpretation?

(a) Apply the standard Cox model to the data.

(b) The analysis in (a) is incorrect as it ignores a possible dependence between observations from the same person. Now take `id`-information into account by applying the marginal Cox model. Use `coxph` with the `+cluster(id)` option in the formula.

(c) Formulate and fit the stratified Cox model where you use `id` as stratification variable. Use `coxph` with the `+strata(id)` option in the formula. What do you think about this model?

(d) Fit a conditional proportional hazards shared gamma frailty model to the data. Use `coxph` with the `+frailty(id)` option in the formula. You can also use `fitfrail` from the package `frailtySurv` with `frailty="gamma"`. In this case, extract the variance with the command `vcov`.

(e) Fit a conditional proportional hazards shared log-normal frailty model to the data. Use either the function `coxme` from the library with the same name with the option `+(1|id)` or `fitfrail` from the package `frailtySurv` with `frailty="lognormal"`. Is there a substantial clustering effect?

Exercise 2 : Intensities and filtrations

Consider a model with marginal intensity $\lambda_{ij}^{\mathcal{F}^{ij}}(t)$ (eg. a Cox model).

(a) Respecting the marginal intensity, construct a random effects model with conditonal intensity

$$Z_i \lambda_{ij}^*(t), \tag{1}$$

where Z_i is a cluster-shared random effect with Laplace transform $\phi(u) = E[\exp(-uZ)]$. That is, find the function $\lambda_{ij}^*(t)$ such that (1) has intensity $\lambda_{ij}^{\mathcal{F}^{ij}}$ with respect to the marginal filtration.

(b) Use your result from (a) above to establish that the observed intensity $\lambda_{ij}^{\mathcal{F}}$ of the model is

$$\begin{aligned}\lambda_{ij}^{\mathcal{F}}(t) = \lambda_{ij}^{\mathcal{F}^{ij}}(t) \exp\left(-\Lambda_{ij}^{\mathcal{F}^{ij}}(t)\right) \left(D\phi_\theta^{-1}\right) \left(\exp\left(-\Lambda_{ij}^{\mathcal{F}^{ij}}(t)\right)\right) \\ \times \frac{\left(D^{1+\sum_{j=1}^k N_{ij}(t)}\right) \phi_\theta \left(\sum_{j=1}^k \phi_\theta^{-1}\left(\exp\left(-\Lambda_{ij}^{\mathcal{F}^{ij}}(t)\right)\right)\right)}{\left(D^{\sum_{j=1}^k N_{ij}(t)}\right) \phi_\theta \left(\sum_{j=1}^k \phi_\theta^{-1}\left(\exp\left(-\Lambda_{ij}^{\mathcal{F}^{ij}}(t)\right)\right)\right)},\end{aligned}$$

where $\Lambda_{ij}^{\mathcal{F}^{ij}}(t) = \int_0^t \lambda_{ij}^{\mathcal{F}^{ij}}(s)ds$.

Hint: You may want to use that

$$\begin{aligned}Y_{ij}(t) E\left(Z_i \middle| \mathcal{F}_{t-}^{ij}\right) &= -Y_{ij}(t)(D\log\phi)\left(\Lambda_{ij}^*(t)\right) \\ E\left(Z_i \middle| \mathcal{F}_{t-}\right) &= -\frac{D^{1+\sum_j N_{ij}(t)}\phi\left(\sum_j \Lambda_{ij}^*(t)\right)}{D^{\sum_j N_{ij}(t)}\phi\left(\sum_j \Lambda_{ij}^*(t)\right)}\end{aligned}$$

where $\Lambda_{ij}^*(t) = \int_0^t \lambda_{ij}^*(s)ds$.

Exercise 3 : Two-stage modelling

Consider the dataset `mena` from the `mets` package.

```
library(mets)
data(mena)
```

The dataset contains simulated menarche ages (`agemena`) for twin pairs (`id`).

(a) Estimate the marginal effect of cohort (`cohort`) and assess wheter the zygosity (`zyg`) affects the marginal models.

(b) Fit separately marginal Cox models for MZ and DZ twins. Fit Clayton-Oakes models by the two-stage method and report Kendall's τ. You may fit the marginal and two-stage models with the function `cox.aalen` and `two.stage` from the `timereg` package.

(c) The `two.stage` function allows the frailty variance to depend on cluster level covariates. Formally compare (test) the strength of the dependence between mono- and dizygotic twins. *Hint:* Use the argument `theta.des` to the `two.stage` function. See the documentation for an example comparing the variance among the two strata in the diabetes data from Exercise 1.

(d) How does your Kendall's τ estimates in (b) and (c) compare to those from the standard frailty model?

Exercise 4 : Intensities in the shared gamma frailty Cox model

Consider again the twin data from Exercise 3.

(a) Fit separately marginal Cox models for MZ and DZ twins. Plot the cumulative marginal (baseline) intensities and compare them. *Hint:* The cumulative baseline can be extracted from the `coxph` object by the function `basehaz(.,centered=FALSE)`.

(b) Fit separately Cox shared gamma frailty models to the MZ and DZ twins. Add the cumulative conditional (basic) intensities to the plot and compare them to the marginal cumulative intensities. To estimate the basic hazard you can use `fitfrail` from the package `frailtySurv` with `frailty="gamma"`. The basic hazard can be extracted with the function `Lambda.fun` from the `fitfrail`-object.

(c) Use the expressions from the lecture to calculate the marginal intensities from the conditional (basic) intensities. Draw these in the plot and compare.

(d) How can you make illustrate the survival for MZ and DZ twins? Simulate a number of frailties and plot the survival curves for these realizations of the frailty.