Recurrent events with R (Part IV) - Joint models

Juan R Gonzalez

Contents

1 Introduction

Objectives

- Understand the concept of *joint modelling* in survival analysis with recurrent event data
- Learn how to perform survival analysis with recurrent event data extending the Frailty models to accomodate that the terminal event (censoring) can be related with the event of interest
- Peform data analyses where the scientific question is to determine factors associated with time until re-ocurrences of a repeated event where the censoring process is informative (e.g terminal event). The model also allow to deal with different covariates and addressed the heterogeneity accross individuals by using frailties.

2 Joint Frailty model

In some occassions the time frame for an individual's repeated hospital admission process may depend on a terminating event, such as death. Therefore, the assumption of noninformative censoring of the recurrent event process by terminating event, can be violated. This dependence should be accounted for in the joint modelling of recurrent events and terminal events. Rondeau et al., 2007 proposed a non-parametric penalized likelihood method for estimating hazard functions in a general joint frailty model for recurrent events and terminal events with right censored survival data. The joint model for the hazard functions for recurrent event ($r_i(\dot{})$) and terminal event ($\lambda_i(\cdot)$) is:

$$\begin{cases} r_i(t|, w_i) = w_i r_0(t) exp(\beta_1' Z_i(t)) = w_i r_i(t) \\ \lambda_i(t|, w_i) = w_i^{\alpha} \lambda_0(t) exp(\beta_2' Z_i(t)) = w_i^{\alpha} \lambda_i(t) \end{cases}$$

where the effect of the explanatory variables is assumed to be different for recurrent event and for death times. w_i are frailties and again the frailty density is assumed that follows a gamma distribution with mean 1 and unknown variance ξ. The previous number of recurrent events can also be considered as an internal time-dependent covariate. The parameter α allows us to quantify the association between the recurrent event and terminal event: if $\alpha = 0$ means that $\lambda_i(t)$ does not depend on w_i and thus death (or the terminal event process) is not informative for the recurrent event rate $r_i(t)$, i.e, the tow rates $\lambda_i(t)$ and $r_i(t)$ are not associated, conditional on covariates; when $\alpha = 1$, the effect of the frailty is identical for the recurrent events and for the terminating event; and when $\alpha > 1$, the recurrent rate and the death rate are positively associated: higher frailty will result in higher risk of recurrence and higher risk of death.

We can fit that model using the function `frailtyPenal` into the library `frailtypack`. We use *readmission* data set to illustrate how to fit this model:

```
data(readmission, package="frailtypack")
head(readmission)
##   id enum t.start t.stop time event      chemo    sex dukes charlson death
## 1  1    1       0     24   24     1    Treated Female     D        3     0
## 2  1    2      24    457  433     1    Treated Female     D        0     0
## 3  1    3     457   1037  580     0    Treated Female     D        0     0
## 4  2    1       0    489  489     1 NonTreated   Male     C        0     0
## 5  2    2     489   1182  693     0 NonTreated   Male     C        0     0
## 6  3    1       0     15   15     1 NonTreated   Male     C        3     0
```

The model assuming gap time formulation is fitted by

```
library(frailtypack)
modJoint.gap <- frailtyPenal(Surv(time,event) ~ cluster(id) + sex +
                                   dukes + charlson + terminal(death),
                             formula.terminalEvent=~sex+dukes+charlson,
                             data=readmission, n.knots=14,
                             kappa=c(9.55e+9,1.41e+12),
                             recurrentAG=FALSE)
##
## Be patient. The program is computing ...
## The program took 429.7 seconds
modJoint.gap
## Call:
## frailtyPenal(formula = Surv(time, event) ~ cluster(id) + sex +
##     dukes + charlson + terminal(death), formula.terminalEvent = ~sex +
##     dukes + charlson, data = readmission, recurrentAG = FALSE,
##     n.knots = 14, kappa = c(9.55e+09, 1.41e+12))
##
##
##   Joint gamma frailty model for recurrent and a terminal event processes
##   using a Penalized Likelihood on the hazard function
##
## Recurrences:
## -------------
##                  coef exp(coef) SE coef (H) SE coef (HIH)        z
## sexFemale   -0.510370  0.600274    0.139736      0.139736 -3.65238
## dukesC       0.406599  1.501701    0.153835      0.153835  2.64309
## dukesD       1.269434  3.558837    0.199200      0.199200  6.37265
## charlson1-2  0.393195  1.481708    0.255267      0.255267  1.54033
## charlson3    0.429569  1.536596    0.136374      0.136374  3.14994
##                      p
## sexFemale   2.5982e-04
## dukesC      8.2153e-03
## dukesD      1.8579e-10
## charlson1-2 1.2348e-01
## charlson3   1.6331e-03
##
##             chisq df global p
## dukes    47.5966  2 4.62e-11
```

```
## charlson 12.2947  2 2.14e-03
##
## Terminal event:
## ----------------
##                  coef exp(coef) SE coef (H) SE coef (HIH)        z
## sexFemale   -0.325843  0.721918    0.219337      0.219337 -1.48559
## dukesC       0.917163  2.502181    0.333981      0.333981  2.74615
## dukesD       2.731926 15.362452    0.376075      0.376075  7.26431
## charlson1-2  0.724711  2.064134    0.623116      0.623116  1.16304
## charlson3    1.114735  3.048759    0.245893      0.245893  4.53341
##                     p
## sexFemale   1.3739e-01
## dukesC      6.0299e-03
## dukesD      3.7492e-13
## charlson1-2 2.4481e-01
## charlson3   5.8040e-06
##
##           chisq df global p
## dukes    63.843  2 1.37e-14
## charlson 20.849  2 2.97e-05
##
##  Frailty parameters:
##    theta (variance of Frailties, w): 0.725101 (SE (H): 0.105399 ) p = 3.0014e-12
##    alpha (w^alpha for terminal event): 0.736805 (SE (H): 0.220026 ) p = 0.00081187
##
##    penalized marginal log-likelihood = -4133.16
##    Convergence criteria:
##    parameters = 2.05e-05 likelihood = 2.45e-05 gradient = 1.78e-07
##
##    LCV = the approximate likelihood cross-validation criterion
##          in the semi parametric case     = 4.85082
##
##    n observations= 861  n subjects= 403
##    n recurrent events= 458
##    n terminal events= 109
##    n censored events= 403
##    number of iterations:  17
##
##    Exact number of knots used:  14
##    Value of the smoothing parameters:  9.55e+09 1.41e+12
```

Here we observe that the terminal event can also depend on other covariates, that's the reason why the argument `formula.terminalEvent` is introduced. The time scale can be changed to calendar time by executing:

```
modJoint.calendar <- frailtyPenal(Surv(t.start,t.stop,event) ~
                                    cluster(id) + sex + dukes +
                                    charlson + terminal(death),
                                    formula.terminalEvent = ~ sex +
                                    dukes + charlson,
                                    data=readmission, n.knots=10,
                                    kappa=c(9.55e9,1.41e12),
                                    recurrentAG=TRUE)
##
## Be patient. The program is computing ...
```

```
## The program took 24.3 seconds
modJoint.calendar
## Call:
## frailtyPenal(formula = Surv(t.start, t.stop, event) ~ cluster(id) +
##     sex + dukes + charlson + terminal(death), formula.terminalEvent = ~sex +
##     dukes + charlson, data = readmission, recurrentAG = TRUE,
##     n.knots = 10, kappa = c(9.55e+09, 1.41e+12))
##
##
##   Joint gamma frailty model for recurrent and a terminal event processes
##   using a Penalized Likelihood on the hazard function
##
## Recurrences:
## -------------
##                  coef exp(coef) SE coef (H) SE coef (HIH)        z
## sexFemale   -0.595825  0.551108    0.151346      0.151346 -3.93684
## dukesC       0.426623  1.532075    0.166207      0.166207  2.56682
## dukesD       1.566323  4.789007    0.220314      0.220314  7.10949
## charlson1-2  0.487908  1.628905    0.280226      0.280226  1.74112
## charlson3    0.592548  1.808591    0.146592      0.146592  4.04217
##                      p
## sexFemale   8.2562e-05
## dukesC      1.0264e-02
## dukesD      1.1647e-12
## charlson1-2 8.1662e-02
## charlson3   5.2959e-05
##
##            chisq df global p
## dukes    57.1334  2 3.92e-13
## charlson 19.3706  2 6.22e-05
##
## Terminal event:
## ----------------
##                  coef exp(coef) SE coef (H) SE coef (HIH)        z
## sexFemale   -0.342743  0.709821    0.216140      0.216140 -1.58574
## dukesC       0.909544  2.483189    0.327533      0.327533  2.77696
## dukesD       2.691115 14.748110    0.362287      0.362287  7.42812
## charlson1-2  0.716614  2.047488    0.617509      0.617509  1.16049
## charlson3    1.119702  3.063940    0.243920      0.243920  4.59044
##                      p
## sexFemale   1.1280e-01
## dukesC      5.4870e-03
## dukesD      1.1013e-13
## charlson1-2 2.4585e-01
## charlson3   4.4230e-06
##
##            chisq df global p
## dukes    67.5611  2 2.11e-15
## charlson 21.3524  2 2.31e-05
##
##  Frailty parameters:
##    theta (variance of Frailties, w): 0.976134 (SE (H): 0.0963049 ) p = 0
##    alpha (w^alpha for terminal event): 0.574631 (SE (H): 0.14728 ) p = 9.5551e-05
##
```

```
##    penalized marginal log-likelihood = -4202.93
##    Convergence criteria:
##    parameters = 3.89e-05 likelihood = 6.95e-05 gradient = 9.11e-09
##
##    LCV = the approximate likelihood cross-validation criterion
##          in the semi parametric case     = 4.92238
##
##    n observations= 861  n subjects= 403
##    n recurrent events= 458
##    n terminal events= 109
##    n censored events= 403
##    number of iterations:  15
##
##    Exact number of knots used:  10
##    Value of the smoothing parameters:  9.55e+09 1.41e+12
```

Notice that in that case, the argument `recurrentAG` must be set to `TRUE`. Log-normal distribution of frailties can be assumed. The model in that case is the same just changing the argument `RandDist` to `LogN`:

```
modJoint.log <- frailtyPenal(Surv(t.start,t.stop,event) ~ cluster(id) +
                              sex + dukes + charlson +
                              terminal(death),
                              formula.terminalEvent = ~ sex + dukes +
                                                        charlson,
                              data=readmission, n.knots=10,
                              kappa=c(9.55e9,1.41e12),
                              recurrentAG=TRUE, RandDist="LogN")
##
## Be patient. The program is computing ...
## The program took 18.11 seconds
modJoint.log
## Call:
## frailtyPenal(formula = Surv(t.start, t.stop, event) ~ cluster(id) +
##     sex + dukes + charlson + terminal(death), formula.terminalEvent = ~sex +
##     dukes + charlson, data = readmission, recurrentAG = TRUE,
##     n.knots = 10, kappa = c(9.55e+09, 1.41e+12), RandDist = "LogN")
##
##
##   Joint Log-Normal frailty model for recurrent and a terminal event processes
##   using a Penalized Likelihood on the hazard function
##
## Recurrences:
## -------------
##                   coef exp(coef) SE coef (H) SE coef (HIH)        z
## sexFemale   -0.530514  0.588303    0.168643      0.168643 -3.14577
## dukesC       0.471227  1.601959    0.189974      0.189974  2.48048
## dukesD       1.794180  6.014540    0.235014      0.235014  7.63436
## charlson1-2  0.348188  1.416499    0.303897      0.303897  1.14574
## charlson3    0.492214  1.635935    0.151156      0.151156  3.25634
##                       p
## sexFemale   1.6565e-03
## dukesC      1.3121e-02
## dukesD      2.2649e-14
## charlson1-2 2.5190e-01
```

```
## charlson3   1.1286e-03
##
##            chisq df global p
## dukes    64.4363  2 1.02e-14
## charlson 11.9165  2 2.58e-03
##
## Terminal event:
## ----------------
##                   coef exp(coef) SE coef (H) SE coef (HIH)        z
## sexFemale   -0.316606  0.728618    0.219481      0.219481 -1.44252
## dukesC       0.928307  2.530223    0.338170      0.338170  2.74509
## dukesD       2.832872 16.994194    0.381173      0.381173  7.43199
## charlson1-2  0.686666  1.987079    0.621854      0.621854  1.10422
## charlson3    1.044942  2.843234    0.244781      0.244781  4.26888
##                       p
## sexFemale   1.4916e-01
## dukesC      6.0495e-03
## dukesD      1.0703e-13
## charlson1-2 2.6950e-01
## charlson3   1.9646e-05
##
##            chisq df global p
## dukes    68.7882  2 1.11e-15
## charlson 18.4844  2 9.69e-05
##
##  Frailty parameters:
##    sigma square (variance of Frailties, eta): 1.25284 (SE (H): 0.198538 ) p = 1.3921e-10
##    alpha (exp(alpha.eta) for terminal event): 0.601129 (SE (H): 0.13567 ) p = 9.3883e-06
##
##    penalized marginal log-likelihood = -4197.74
##    Convergence criteria:
##    parameters = 3.16e-05 likelihood = 7.36e-05 gradient = 4.34e-09
##
##    LCV = the approximate likelihood cross-validation criterion
##          in the semi parametric case     = 4.91658
##
##    n observations= 861  n subjects= 403
##    n recurrent events= 458
##    n terminal events= 109
##    n censored events= 403
##    number of iterations:  13
##
##    Exact number of knots used:  10
##    Value of the smoothing parameters:  9.55e+09 1.41e+12
```

3 Recommended lectures

In the GitHub folder corresponding to this lecture there is a paper describing how a real data set is analyzed (read Sections 5.2 and 5.3 of the file Joint_Frailty_model_application.pdf). The file frailtypack paper (Rondeau, Mazroui and Gonzalez, 2012) that is available in the material of Sesion 7, describes how to fit these models using `frailtypack` package.

4 Exercise (to deliver)

The folder `https://github.com/isglobal-brge/TeachingMaterials/tree/master/Longitudinal_data_analysis/data` contains the `copd_recurrent.txt` file encoding hospital readmission of patients diagnosed with chronic obstructive pulmonary disease (COPD). The variable *id* encodes the unique patient identificator (to be used for clustering the data). The researchers are interested in studing the effect of physical activity (variable *phys.act*) with regard to the probability of being hospital readmitted (variables *time.readmission* and *status.readmission*. They know that pulmonary capacity (variable *fev*), smoking (variable *smoke*) and age (variable *age*) also affect the likelihood of coming back to the hospital. In addition, they also have information about the terminal event (variables *time.death* and *status.death*).

Exercise:

Analyze the data by using AG model, frailty model and joint model to determine whether physical activity changes the probability of being hospital readmitted. Compare the results and provide a biomedical conclusion.

---

5 References

- The [`frailtypack`] package (https://cran.r-project.org/web/packages/frailtypack/)
- V. Rondeau, S. Mathoulin-Pellissier, H. Jacqmin-Gadda, V. Brouste, P. Soubeyran (2007). Joint frailty models for recurring events and death using maximum penalized likelihood estimation:application on cancer events. Biostatistics 8,4, 708-721.
- Y. Mazroui, S. Mathoulin-Pelissier, P. Soubeyranb, V. Rondeau (2012) General joint frailty model for recurrent event data with a dependent terminalevent: Application to follicular lymphoma data. Statistics in Medicine, 31, 11-12, 1162-1176.
- V. Rondeau, Y. Mazroui and J. R. Gonzalez (2012). Frailtypack: An R package for the analysis of correlated survival data with frailty models using penalized likelihood estimation or parametric estimation. Journal of Statistical Software 47, 1-28.

6 Session information

```
## R version 3.4.1 (2017-06-30)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 16299)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=Spanish_Spain.1252  LC_CTYPE=Spanish_Spain.1252
## [3] LC_MONETARY=Spanish_Spain.1252 LC_NUMERIC=C
## [5] LC_TIME=Spanish_Spain.1252
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] frailtypack_2.12.6 survC1_1.0-2       MASS_7.3-47
## [4] boot_1.3-19        survival_2.41-3    knitr_1.20
## [7] BiocStyle_2.4.1
```

```
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_0.12.12    lattice_0.20-35 digest_0.6.12   rprojroot_1.3-2
##  [5] grid_3.4.1      nlme_3.1-131    backports_1.1.0 magrittr_1.5
##  [9] evaluate_0.10.1 stringi_1.1.6   Matrix_1.2-10   rmarkdown_1.8
## [13] splines_3.4.1   statmod_1.4.30  tools_3.4.1     stringr_1.3.0
## [17] yaml_2.1.16     compiler_3.4.1  htmltools_0.3.6
```