Fitting joint model for longitudinal ordinal and multistate process in INLA

Elsa Vazquez Arreola

2023-08-02

You will need packages *mstate, haven, INLAjoint, and INLA.*

```
#read packages necessary to run this analysis
library(mstate)

## Loading required package: survival

library(haven)

## Warning: package 'haven' was built under R version 4.2.2

library(INLA)

library(INLAjoint)
```

The first step is to obtain data for multistate model in the right format. For that we use the *mstate* package in R and its functions.

```
#converting imported dataset into a data frame in R to avoid errors later
ms_cvd_mortality=as.data.frame(ms_cvd_mortality)



#defining matrix for state transitions
tmat <- matrix(NA, 3, 3)
tmat[1, 2:3] <- 1:2
tmat[2, 3] <- 3
tmat

##      [,1] [,2] [,3]
## [1,]   NA    1    2
## [2,]   NA   NA    3
## [3,]   NA   NA   NA

#the dimension names of our matrix will be healthy, nonfatal CVD and death
 dimnames(tmat) <- list(from = c("healthy", "nonfatal CVD", "CVD death"), to
= c("healthy",
                  "nonfatal CVD", "CVD death"))
tmat
```

```
##               to
## from           healthy nonfatal CVD CVD death
##   healthy           NA            1         2
##   nonfatal CVD      NA           NA         3
##   CVD death         NA           NA        NA
```

```
#defining transition matrix
tmat <- transMat(x = list(c(2, 3), c(3), c()), names = c("healthy", "nonfatal
CVD", "CVD death"))
tmat
```

```
##               to
## from           healthy nonfatal CVD CVD death
##   healthy           NA            1         2
##   nonfatal CVD      NA           NA         3
##   CVD death         NA           NA        NA
```

```
#specifying that transitions are for illness to death model
tmat <- trans.illdeath(names = c("healthy", "nonfatal CVD", "CVD death"))
tmat
```

```
##               to
## from           healthy nonfatal CVD CVD death
##   healthy           NA            1         2
##   nonfatal CVD      NA           NA         3
##   CVD death         NA           NA        NA
```

```
paths(tmat)
```

```
##      [,1] [,2] [,3]
## [1,]    1   NA   NA
## [2,]    1    2   NA
## [3,]    1    2    3
## [4,]    1    3   NA
```

```
#getting data for multistate model
covs <- c("female", "ili", "takes_ins", "Diab_Dur", "cvd_his", "bmi_bas", "bl
ack", "hispan", "other_race","native",
          "b_age10")

ms_cvdmortality <- msprep(time = c(NA, "fu_nonfatal_cvd_years", "fu_cvd_death
_years"), status = c(NA,"nonfatal_cvd_event",

"CVD_death"),
                      data = ms_cvd_mortality, trans = tmat, keep = covs, id
="sid")


#checking how many subjects transitioned into each of the states
events(ms_cvdmortality) #do not take out other race and native from any trans
ition
```

```
## $Frequencies
##               to
## from           healthy nonfatal CVD CVD death no event total entering
##   healthy            0          708        65     4081           4854
##   nonfatal CVD       0            0        34      674            708
##   CVD death          0            0         0       99             99
##
## $Proportions
##               to
## from              healthy nonfatal CVD  CVD death   no event
##   healthy      0.00000000   0.14585909 0.01339102 0.84074990
##   nonfatal CVD 0.00000000   0.00000000 0.04802260 0.95197740
##   CVD death    0.00000000   0.00000000 0.00000000 1.00000000

######  ID variable must be converted to an integer

ms_cvdmortality$id=as.integer(ms_cvdmortality$sid)
```

More information about the multistate package and its functions and uses can be found in *de Wreede LC., Fiocco M, Putter H. mstate: An R package for the analysis of competing risks and Multi-State Models. Journal of Statistical Software; 2011, 38(7).*

```
###### separate data for each transition, this will be necessary later, when specifying model in INLA
transition1=ms_cvdmortality[ms_cvdmortality$trans==1, ]
transition2=ms_cvdmortality[ms_cvdmortality$trans==2, ]
transition3=ms_cvdmortality[ms_cvdmortality$trans==3, ]
```

The following steps should be followed to get data into a format accepted by *INLA and INLAjoint* packages. These steps are closely following instructions for the *INLAjoint* package that can be found at https://github.com/DenisRustand/INLAjoint/blob/main/vignettes/INLAjoint.pdf

```
################################################################################
#############  GETTING DATA FOR LONGITUDINAL PART OF THE MODEL  ################
################################################################################
#import longitudinal dataset
longitudinal_data <- read_sas("C:/longitudinal_data.sas7bdat",
NULL)

#transform dataset from list to data frame in R
longitudinal_mortalitys=as.data.frame(longitudinal_data)

typeof(longitudinal_mortalitys$sid)
```

```
## [1] "double"

#If ID variable is not of "integer" type, transform it to integer

longitudinal_mortalitys$id=as.integer(longitudinal_mortalitys$sid)

typeof(longitudinal_mortalitys$id)

## [1] "integer"

#keep only observations without missing data and only variables that will be
used to fit the model for the time-dependent covariate

Longi <- na.omit(longitudinal_mortalitys[, c("female", "ili",
"cvd_his", "time_years2", "ckd_risk_num","id", "bmi_bas", "b_age10")])

typeof(Longi$id)

## [1] "integer"

##############################################################################
## joint model for ordinal risk catefory of CKD and CVD mortality process  ##
##############################################################################

library(INLA)
library(INLAjoint)


#setting up multistate outcomes in INLA for each transition
t1=inla.surv(time=transition1$Tstop, event=transition1$status)
t2=inla.surv(time=transition2$Tstop, event=transition2$status)
t3=inla.surv(time=transition3$Tstop, truncation = transition3$Tstart, event =
transition3$status)
```

To fit the backward continuation ratio logistic regression model, we have to transform the data into the format explained at *Bender R, Benner A. Calculating ordinal regression models in SAS and S-plus. Biometrical Journal; 2000, 6:677-699*.

The JMbayes2 package in R contains the function cr_setup, which transforms the data into the right format to fit the backward continuation ratio logistic regression model. More information on how to use this function can be found at https://rdrr.io/github/drizopoulos/JMbayes2/man/cr_setup.html

```
##############################################################################
############## setting data up for continuation ratio model ##################
##############################################################################

library(JMbayes2)

cr_vals <- cr_setup(Longi$ckd_risk_num, "backward")
cr_data <- Longi[cr_vals$subs, ]
cr_data$y_new <- cr_vals$y
cr_data$cohort <- cr_vals$cohort


#create dummy variables for each category to fit model
cr_data$cat2=ifelse(cr_data$cohort=="y<=2", 1, 0)
cr_data$cat3=ifelse(cr_data$cohort=="y<=3", 1, 0)
cr_data$cat4=ifelse(cr_data$cohort=="all", 1,0)
cr_data$treatment=ifelse(cr_data$ili=="1", 1, 0)


##### fitting joint model
mjoint2=joint(formSurv = list(t1~ili+female+b_age10+bmi_bas+cvd_his,
                              t2~ili+female+b_age10+bmi_bas+cvd_his,
                              t3~ili+female+b_age10+bmi_bas+cvd_his),
              formLong = list(y_new ~ -1+cat2+cat3+cat4+treatment+female+b_ag
e10+bmi_bas+time_years2+
                                treatment*time_years2+(1+time_years2|id)),
              basRisk = c("rw2", "rw2", "rw2"),
              dataSurv = list(transition1, transition2, transition3),
              dataLong = cr_data, id="id", timeVar = "time_years2", corLong =
FALSE,
              family=c("binomial"),
              assoc=list(c("CV", "CV", "CV")),
              control=list(verbose=TRUE,strategy="adaptive", int.strategy="eb
", priorFixed=list(mean=0, prec=0.16,

mean.intercept=0, prec.intercept=0.16),
                           priorAssoc=list(mean=0, prec=0.16) ))
```

```
##### exploring results of joint model

summary(mjoint2, hazr=TRUE)

## Longitudinal outcome (binomial)
##                               mean     sd 0.025quant 0.5quant 0.975quant
## cat2_L1                    -8.5386 0.4626    -9.4453  -8.5386    -7.6318
## cat3_L1                   -11.5942 0.4655   -12.5065 -11.5942   -10.6819
## cat4_L1                   -13.9468 0.4700   -14.8680 -13.9468   -13.0256
## treatment_L1               -0.2606 0.0901    -0.4372  -0.2606    -0.0840
## female_L1                  -0.3061 0.0827    -0.4682  -0.3061    -0.1440
## b_age10_L1                  0.6932 0.0582     0.5791   0.6932     0.8073
## bmi_bas_L1                  0.0671 0.0068     0.0538   0.0671     0.0804
## time_years2_L1              0.2061 0.0091     0.1883   0.2061     0.2239
## treatment:time_years2_L1   -0.0106 0.0129    -0.0359  -0.0106     0.0146
##
## Random effects variance-covariance (L1)
##                                 mean     sd 0.025quant 0.5quant 0.975quant
## Intercept_L1                  5.7331 0.3567     5.1606   5.6888     6.5459
## time_years2_L1                0.0528 0.0099     0.0380   0.0512     0.0768
## Intercept_L1:time_years2_L1 0.0053 0.0446    -0.0602  -0.0023     0.1147
##
## Survival outcome (S1)
##                             exp(mean)     sd 0.025quant 0.5quant 0.975quant
## Baseline risk (variance)_S1    0.0260 0.0138     0.0122   0.0218     0.0642
## Baseline risk (mean)_S1        0.0047 0.0022     0.0017   0.0042     0.0103
## ili1_S1                        0.9548 0.0708     0.8230   0.9520     1.1013
## female_S1                      0.6070 0.0471     0.5196   0.6051     0.7048
## b_age10_S1                     1.3307 0.0750     1.1892   1.3284     1.4839
## bmi_bas_S1                     0.9945 0.0066     0.9816   0.9945     1.0076
## cvd_his_S1                     3.6282 0.2939     3.0843   3.6159     4.2390
##
## Survival outcome (S2)
##                             exp(mean)     sd 0.025quant 0.5quant 0.975quant
## Baseline risk (variance)_S2    0.0148 0.0033     0.0091   0.0146     0.0221
## Baseline risk (mean)_S2        0.0017 0.0030     0.0000   0.0008     0.0098
## ili1_S2                        1.2906 0.3161     0.7783   1.2529     2.0166
## female_S2                      0.5848 0.1508     0.3429   0.5659     0.9337
## b_age10_S2                     1.3036 0.2162     0.9295   1.2857     1.7782
## bmi_bas_S2                     0.9874 0.0205     0.9478   0.9872     1.0282
## cvd_his_S2                     3.6449 0.9616     2.1106   3.5221     5.8762
##
## Survival outcome (S3)
##                             exp(mean)     sd 0.025quant 0.5quant 0.975quant
## Baseline risk (variance)_S3    0.0139 0.0029     0.0085   0.0139     0.0199
## Baseline risk (mean)_S3        0.0585 0.1571    -0.0002   0.0131     0.4228
## ili1_S3                        0.5532 0.2057     0.2537   0.5177     1.0560
## female_S3                      0.2422 0.1293     0.0782   0.2128     0.5779
## b_age10_S3                     1.3434 0.2993     0.8493   1.3107     2.0225
## bmi_bas_S3                     0.9495 0.0304     0.8910   0.9489     1.0106
## cvd_his_S3                     2.4837 0.9202     1.1424   2.3256     4.7318
##
```

```
## Association longitudinal - survival
##            mean     sd 0.025quant 0.5quant 0.975quant
## CV_L1_S1 0.0993 0.0264     0.0494   0.0986     0.1530
## CV_L1_S2 0.3621 0.1070     0.1796   0.3519     0.5950
## CV_L1_S3 0.4813 0.0839     0.3253   0.4778     0.6545
##
## log marginal-likelihood (integration)    log marginal-likelihood (Gaussian)
##                              -273779.0                              -273770.4
##
## Deviance Information Criterion:  -483386.2
## Widely applicable Bayesian information criterion:  -504411.2
## Computation time: 2235.16 seconds
```

To estimate time-dependent association of logit for probability of being in a higher risk category relative to lower risk categories of CKD progression with transition 1, transition 2, and transition 3, exponentiate mean of CV_L1_S1, CV_L1_S2, and CV_L1_S3, respectively.