Analysis of multivariate survival data based on Case Control Data

Klaus Holst & Thomas Scheike

August 5, 2019

Overview

When looking at multivariate survival data with the aim of learning about the dependence that is present, possibly after correcting for some covariates different approaches are available in the mets package

- Binary models and adjust for censoring with inverse probabilty of censoring weighting
  - biprobit model
- Bivariate surival models of Clayton-Oakes type
  - With regression structure on dependence parameter
  - With additive gamma distributed random effects
  - Special functionality for polygenic random effects modelling such as ACE, ADE ,AE and so forth.
- Plackett OR model model
  - With regression structure on OR dependence parameter
- Cluster stratified Cox

We have discussed how to fit such models in the vignette about twostage survival modelling. Here we show what can be done if one has data available from case-control sampling.

First we set up some case-control data

```
library(mets)
set.seed(100)
ncases <- 2000
ncontrols <- ncases*5
data <- simClaytonOakes.twin.ace(100000,1,2,0,3,Cvar=1)
theta <- c(1,2)
cens.prob <- mean(data$status==0)
#
data2 <- fast.reshape(data,id="cluster")
with(data2,table(status1,status2))

controls <- which(data2$status2==0)
cases <- which(data2$status2==1)
cases <- sample(cases,min(ncases,length(cases)))
controls <- sample(controls,min(ncontrols,length(controls)))
nccc <- c(length(cases),length(controls))
```

```
clustco <- data2$cluster[controls]
clustca <- data2$cluster[cases]
#
med <- data$cluster %in% c(clustco,clustca)
datacc <- data[med,]
datacc2 <- fast.reshape(datacc,id="cluster")
dd <- with(datacc2,table(status1,status2))
#
#
out <- twin.polygen.design(data,id="cluster")
pardes <- out$pardes
des.rv <- out$des.rv

aa <- phreg(Surv(time,status)~+cluster(cluster),data=data)

out <- twin.polygen.design(datacc,id="cluster")
pardes <- out$pardes
des.rv <- out$des.rv
#
#
# needs to use pair structure to profile out
# baseline
mm <- familycluster.index(datacc$cluster)
pairs <- matrix(mm$familypairindex,ncol=2,byrow=TRUE)
#
kinship <- rep(1,nrow(pairs))
kinship[datacc$zyg[pairs[,1]]=="DZ"] <- 0.5
table(kinship)
#

dout <- make.pairwise.design(pairs,kinship,type="ace")
des.rv <- dout$random.design
pardes <- dout$theta.des
#
cr.models <- list(Surv(time,status)~+1)
tscce <- survival.twostage(NULL,data=datacc,
          clusters=datacc$cluster,
          theta=theta,var.link=0,step=1.0,
          random.design=des.rv,theta.des=pardes,
          pairs.rvs=dout$ant.rvs,var.par=1, pairs=pairs,
          case.control=1,marginal.status=datacc$status,
          cr.models=cr.models)
summary(tscce)
```

```
Loading required package: timereg
Loading required package: survival
Loading required package: lava
lava version 1.6.3
mets version 1.2.4

Attaching package: 'mets'

The following object is masked _by_ '.GlobalEnv':

    object.defined
       status2
status1     0     1
      0 16121 15661
      1 15828 52390
```

```
kinship
 0.5    1
5963 6037
Dependence parameter for Clayton-Oakes model
Variance of Gamma distributed random effects
$estimates
               Coef.         SE        z P-val Kendall tau         SE
dependence1 1.006966 0.08370828 12.02947     0   0.3348778 0.01851575
dependence2 1.838534 0.08963533 20.51127     0   0.4789678 0.01216686

$type
[1] "clayton.oakes"

$h
            Estimate Std.Err   2.5%  97.5%    P-value
dependence1   0.3539 0.02812 0.2988 0.4090  2.496e-36
dependence2   0.6461 0.02812 0.5910 0.7012 7.193e-117

$vare
NULL

$vartot
   Estimate Std.Err  2.5% 97.5% P-value
p1    2.846 0.06515 2.718 2.973       0

attr(,"class")
[1] "summary.mets.twostage"
```

```
# known baseline from cohort
aa <- aalen(Surv(time,status)~+1,data=data,robust=0)
ts <- survival.twostage(aa,data=datacc,
          clusters=datacc$cluster,
          theta=theta,var.link=0,step=1.0,
          random.design=des.rv,theta.des=pardes,
          pairs.rvs=dout$ant.rvs,var.par=1, pairs=pairs,
          case.control=1,
          marginal.status=datacc$status,
          cr.models=cr.models)
summary(ts)
```

```
Dependence parameter for Clayton-Oakes model
Variance of Gamma distributed random effects
$estimates
               Coef.         SE        z P-val Kendall tau          SE
dependence1 1.032045 0.07944442 12.99078     0   0.3403792 0.017283117
dependence2 1.897001 0.06795064 27.91734     0   0.4867849 0.008948751

$type
[1] "clayton.oakes"

$h
            Estimate Std.Err   2.5%  97.5%    P-value
dependence1   0.3523 0.02247 0.3083 0.3964  2.030e-55
dependence2   0.6477 0.02247 0.6036 0.6917 1.079e-182

$vare
NULL

$vartot
   Estimate Std.Err  2.5% 97.5% P-value
p1    2.929 0.07785 2.776 3.082       0

attr(,"class")
[1] "summary.mets.twostage"
```

Figure **??** shows the baseline

```
plot(aa)
lines(tscce$baseline,col=2)
```

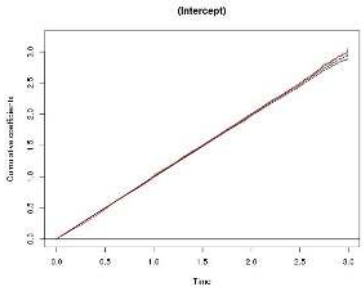

Figure 1: Baseline with robust standard errors. Black based on cohort data, red based on profiling for case-control data.