Marginal modelling of clustered survival data

Klaus Holst & Thomas Scheike

March 3, 2020

Overview

A basic component for our modelling is that all these models are build around marginals that on Cox form. The marginal Cox model can be fitted efficiently in the mets package.

The basic models assumes that each subject has a marginal on Cox-form

$$\lambda_{s(k,i)}(t) \exp(X_{ki}^T \beta).$$

where $s(k,i)$ gives the strata for the subject.

We here discuss the

- robust standard errors of
  - regression parameters
  - baseline
- cumulative residuals score test

First we generate some data from the Clayton-Oakes model, with 5 members in each cluster and a variance parameter at 2

```
library(mets)
set.seed(1000) # to control output in simulatins for
    p-values below.
 n <- 1000
 k <- 5
 theta <- 2
 data <- simClaytonOakes(n,k,theta,0.3,3)
 head(data)
```

```
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
       time status x cluster   mintime lefttime truncated
1 0.1406317      1 0       1 0.1406317        0         0
2 0.4593768      1 0       1 0.1406317        0         0
3 1.0952678      1 0       1 0.1406317        0         0
4 0.2057554      1 1       1 0.1406317        0         0
5 0.6776620      1 0       1 0.1406317        0         0
6 1.6093755      1 0       2 0.1092390        0         0
```

Now fitting the and producing robust standard errors for both regression parameters and baseline.

Note that

$$\hat{A}_s(t) - A_s(t) = \sum_k \sum_i \int_0^t 1/S_s dM_{ki}^s - P^s(t)\beta_k \tag{1}$$

with $P^s(t)$ a derivative wrt to β, and

$$\hat{\beta} - \beta = \sum_k (\sum_i \int_0^\tau (Z_{ik} - E_s) dM_{ik}^s) \tag{2}$$

with

$$M_{ki}(t) = N_{ki}(t) - \int_0^t Y_{ki}(s) \exp(Z_{ki}\beta) d\Lambda_{s(ki)}(t) \tag{3}$$

the basic o-mean processes, that are martingales in the iid setting.

The variance of the baseline of strata s is

$$\sum_k (\sum_i \int_0^t 1/S_{0s(ki)} d\hat{M}_{ki}^s)^2 \tag{4}$$

that can be computed using the particular structure

$$d\hat{M}_{ik}(t) = dN_{ik}(t) - 1/S_{0s(i,k)} \exp(Z_{ik}\beta) dN_{s.}(t) \tag{5}$$

This robust variance of the baseline and the iid decomposition for β is computed in mets as:

```
out <- phreg(Surv(time,status)~x+cluster(cluster),data=data)
summary(out)
# robust standard errors attached to output
rob <- robust.phreg(out)

# making iid decomposition of regression parameters
betaiid <- iid(out)
head(betaiid)
# robust standard errors
crossprod(betaiid)^.5
# same as
```

```
   n events
5000   4854

1000 clusters

  Estimate     S.E.  dU^-1/2 P-value
x 0.287859 0.028177 0.028897       0
           [,1]
1 -3.461601e-04
2 -1.449189e-03
3 -3.898156e-05
4  4.215605e-04
5  3.425390e-04
6 -7.706668e-05
           [,1]
[1,] 0.02817714
```

Looking at the plot with robust standard errors

```
bplot(rob,se=TRUE,robust=TRUE)
```

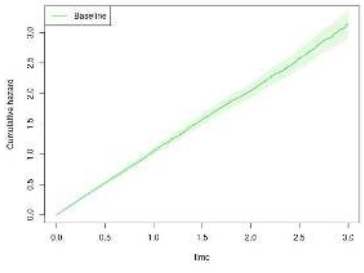

Figure 1: Baseline with robust standard errors.

One can also make survival prediction with robust standard errors using the phreg.

```
pp <- predict(out,data[1:20,],se=TRUE,robust=TRUE)
```

```
plot(pp,se=TRUE,whichx=1:10)
```

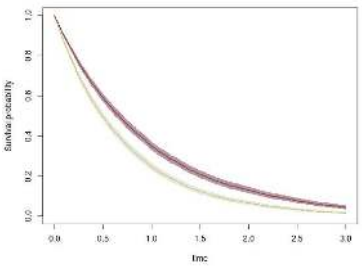

Figure 2: Survival predictions with robust standard errors for Cox model

Finally, just to check that we can recover the model we also estimate the dependence parameter

```
tt <- twostageMLE(out,data=data)
summary(tt)
```

```
Dependence parameter for Clayton-Oakes model
Variance of Gamma distributed random effects
$estimates
                Coef.         SE        z P-val Kendall tau        SE
dependence1 0.5316753 0.03497789 15.20032     0   0.2100093 0.0109146

$type
NULL

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

Goodness of fit

The observed score process is given by

$$U(t,\hat{\beta}) = \sum_k \sum_i \int_0^t (Z_{ki} - \hat{E}_s) d\hat{M}^s_{ki} \tag{6}$$

where s is strata, this has as iid decomposition as

$$\hat{U}(t) = \sum_k \sum_i \int_0^t (Z_{ki} - E_s) dM^s_{ki} - \sum_k I_t \beta_k \tag{7}$$

where β_k is the iid decomposition of the score process for the true β

$$\beta_k = \sum_i \int_0^t (Z_{ki} - E_s) dM^s_{ki} \tag{8}$$

and I_t is the derivative of the total score with respect to β.

This observed score can be resampled given it is on iid form in terms of clusters.

Now using the cumulative score process for checking proportional hazards

```
gout <- gof(out)
gout
```

```
Cumulative score process test for Proportionality:
  Sup|U(t)|  pval
x  30.24353 0.401
```

The p-value reflects wheter the observed score process is consistent with the model.

```
plot(gout)
```

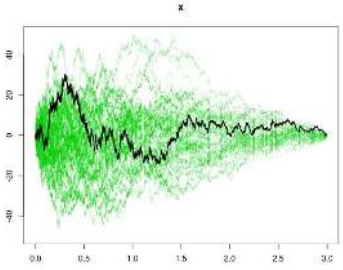

Figure 3: Goodness of fit for clustered Cox model.

Cluster stratified Cox models

For clustered data it is possible to estimate the regression coefficient within clusters by using Cox's partial likelihood stratified on clusters.

Note, here that the data is generated with a different subject specific structure, so we will not recover the β at 0.3 and the model will not be a proportional Cox model, we we would also expect to reject "proportionality" with the gof-test.

The model can be thought of as

$$\lambda_k(t)\exp(X_{ki}^T\beta)$$

where $\lambda_k(t)$ is some cluster specific baseline.

The regression coefficient β can be estimated by using the partial likelihood for clusters.

```
out <- phreg(Surv(time,status)~x+strata(cluster),data=data)
summary(out)
```

```
    n events
 5000   4854

  Estimate     S.E.  dU^-1/2 P-value
x 0.406307 0.032925 0.039226       0
```

The cumulative score processes can still be used to validate the model

```
gg <- gof (out)
summary(gg)
```

```
Cumulative score process test for Proportionality:
  Sup|U(t)|  pval
x  27.55616 0.195
```