Supporting Data Example

In this supporting web material, we have provided a simulated data example for interested readers of *The liability-threshold model for case-control-family studies incorporating the time-aspect* that would like to see the proposed model in action.

The data example has been simulated as in the second simulation study in the article. It mimics data from a case-control-family study on full siblings and half siblings where the aim is to make inference on the cumulative risk of a given disease at age 75 including the role of genetic and environmental factors. Thus, in this application τ is equal to 75 years. There are no covariates included in model.

The variance-covariance matrices of full siblings and half siblings applied in the model are given as

$$\Sigma_{\text{full}} = \begin{pmatrix} \sigma_A^2 + \sigma_C^2 + \sigma_E^2 & \frac{1}{2}\sigma_A^2 + \sigma_C^2 \\ \frac{1}{2}\sigma_A^2 + \sigma_C^2 & \sigma_A^2 + \sigma_C^2 + \sigma_E^2 \end{pmatrix} \quad \text{and} \quad \Sigma_{\text{half}} = \begin{pmatrix} \sigma_A^2 + \sigma_C^2 + \sigma_E^2 & \frac{1}{4}\sigma_A^2 + \sigma_C^2 \\ \frac{1}{4}\sigma_A^2 + \sigma_C^2 & \sigma_A^2 + \sigma_C^2 + \sigma_E^2 \end{pmatrix}.$$

The parameters of interest that will we estimate are:

- the cumulative risk of disease at 75 years $\Phi(\beta_0)$
- the variance of the genetic component σ_A^2
- the variance of the environmental component σ_C^2

With these we can estimate:

- the casewise concordance for full siblings and half siblings
- the relative recurrence risk ratio for full siblings and half siblings

Data

The file `data.csv` contains the data to be analysed. The data are arranged with one *proband–family* member pair per row and include the following variables:

ID The family ID

status1 Status of proband at age τ

- 0: proband is alive and disease-free (control proband)
- 1: proband has experienced the event of interest (case proband)
- 2: proband has experienced the competing event (control proband)

out1 The value of I for proband at age τ

- 0: proband is alive and disease-free at age τ or has experienced the competing event before or at age τ (control proband)
- 1: proband has experienced the event of interest before or at age τ (case proband)

age1 Age of proband

- if status1=0 then age1=τ
- if status1=1 then age1=age at occurrence of the event of interest
- if status1=2 then age1=age at occurrence of the competing event

status2 Status of family member at age τ

- 0: family member is alive and disease-free
- 1: family member has experienced the event of interest
- 2: family member has experienced the competing event
- 3: family member has been right-censored

out2 The value of I for family member at age τ

- 0: family member is alive and disease-free at age τ or has experienced the competing event before or at age τ (control proband) or has been censored before or at age τ (will have IPCW equal to zero in the adjusted model)
- 1: family member has experienced the event of interest before or at age τ

age2 Age of family member

- if status2=0 then age2=τ
- if status2=1 then age2=age at occurrence of the event of interest
- if status2=2 then age2=age at occurrence of the competing event
- if status2=3 then age2=age at right-censoring

rel Family member's relation to proband

- f: full sibling
- h: half sibling

group Grouping of family members according to proband status and family member's relation to proband

- 1: proband is a case proband and the proband and the family member are full siblings
- 2: proband is a case proband and the proband and the family member are half siblings
- 3: proband is a control proband and the proband and the family member are full siblings
- 4: proband is a control proband and the proband and the family member are half siblings

Analysis

Analysis of the data example is done using the provided R files `run.R` and `functions.R`. The file `run.R` is the main file and will fit the IPCW adjusted liability-threshold model employing a composite likelihood to the data and estimate the parameters of interest. The file `functions.R` contains the composite log-likelihood function and the composite score function of the model and is required for analysis.

Here, we will go through the different steps in `run.R`.

First, the required libraries are loaded.

```
library(mets) # Should be version 1.1.1.1 or later
library(numDeriv)
library(mvtnorm)
library(reshape)
```

Then the data example is loaded. As seen, there is one proband-family member pair per row.

```
data <- read.table("data.csv", header=TRUE, sep="\t")
head(data)
```

```
##   ID status1 out1     age1 status2 out2     age2 rel group
## 1  1       0    0 75.00000       0    0 75.00000   f     3
## 2  2       2    0 69.15220       0    0 75.00000   f     3
## 3  3       0    0 75.00000       0    0 75.00000   f     3
## 4  4       2    0 73.76681       3    0 38.69826   f     3
## 5  5       0    0 75.00000       3    0 39.03690   f     3
## 6  6       0    0 75.00000       3    0 38.10148   f     3
```

The file `functions.R` contains the composite log-likelihood function and the composite score function and needs to be loaded.

```
source("functions.R")
```

The IPCWs are estimated using Aalen's additive model and the variable group.

```
# Formula
formula <- Surv(age2, status2==3)~as.factor(group)

# Design matrix
X <- model.matrix(formula, data)

# Fitting Aalen's additive model
fit <- aalen(formula, data, n.sim=0, robust=0)

# Predicting cumulative effects at observed event times
Gcxp <- Cpred(fit$cum, data$age2)[,-1]

# Calculating censoring probabilities at observed event times
Gcx <- exp(-apply(Gcxp*X,1,sum))
data$pc <- Gcx
```

With the censoring probabilities, we can estimate the IPCWs. Note, that if a sibling is censored (status2=3) then the corresponding IPCW is equal to zero.

```
# IPCW
data[,"ipcw"] <- as.numeric(data$status2!=3)/Gcx
head(data)
```

```
##   ID status1 out1     age1 status2 out2     age2 rel group        pc     ipcw
## 1  1       0    0 75.00000       0    0 75.00000   f     3 0.5472660 1.827265
## 2  2       2    0 69.15220       0    0 75.00000   f     3 0.5472660 1.827265
## 3  3       0    0 75.00000       0    0 75.00000   f     3 0.5472660 1.827265
## 4  4       2    0 73.76681       3    0 38.69826   f     3 0.7404955 0.000000
## 5  5       0    0 75.00000       3    0 39.03690   f     3 0.7293508 0.000000
## 6  6       0    0 75.00000       3    0 38.10148   f     3 0.7591956 0.000000
```

Now, we can fit the IPCW adjusted model and estimate the parameters of interest:

```
# Initial values
par <- c(-1,1,1)

# Optimisation
op <- nlminb(par, loglik, gradient=score, outcome="out", data=data, rel="rel", ID="ID",
             weights="ipcw", control=list(trace=0))

# Checking whether model has converged or not
op$convergence # 0: successful convergence

## [1] 0
```

The variance of the parameter estimates are found using the sandwich estimator:

```
# Variance-covariance matrix
H <- jacobian(score, op$par, data=data, outcome="out", ID="ID", rel="rel", weights="ipcw",
              grad=FALSE)
U <- score(op$par, data=data, outcome="out", ID="ID", rel="rel", weights="ipcw", grad=TRUE)
V <- t(U)%*%U
vcv <- solve(H)%*%V%*%solve(H)
```

And finally, we have the results:

```
parest(par=op$par,varcovar=vcv)

##                                              Estimate Std.Err   2.5%  97.5%       P-value
## Cumulative risk at 75 years                    0.1025  0.0029 0.0968 0.1083 1.669515e-266
## sigA^2                                         0.2646  0.1326 0.0046 0.5246  4.605432e-02
## sigC^2                                         0.3637  0.0615 0.2431 0.4843  3.441371e-09
## Casewise concordance full siblings             0.3250  0.0097 0.3059 0.3440 3.040296e-245
## Casewise concordance half siblings             0.2863  0.0173 0.2524 0.3202  1.642493e-61
## Relative recurrence risk ratio full siblings   3.1695  0.1054 2.9629 3.3761 1.344799e-198
## Relative recurrence risk ratio half siblings   2.7924  0.1691 2.4610 3.1237  2.774095e-61
```