Introduction to the R-package rtmle

Thomas Alexander Gerds

Department of Biostatistics, University of Copenhagen

1 About

The R-package `rtmle` implements a class of targeted minimum loss estimators for the analysis of longitudinal register data. The functionality is similar to the established R-package `ltmle` with the following features and focus points:

- explicit handling of competing risks
- polypharmacy
- descriptive statistics
- diagnostic tools (not to be confused with model goodness of fit)
- subgroup analysis
- support for glmnet
- cheap bootstrap inference
- cross-fitting
- parallel computing

2 Data structures

We consider register data in the following format.

```
library(rtmle)
library(prodlim)
library(riskRegression)
library(data.table)
set.seed(118)
ld <- simulate_long_data(n = 10000,number_epochs = 20,beta = list(sum
    _A_on_Y = -.1,A0_on_Y = 0),register_format = TRUE)
ld
```

```
$baseline_data
          id   sex      age
       <int> <int>    <num>
    1:     1     0 59.44409
    2:     2     1 73.01360
    3:     3     0 78.26086
    4:     4     1 66.53408
    5:     5     0 62.35921
   ---
 9996:  9996     0 43.90035
 9997:  9997     1 41.32339
 9998:  9998     0 65.08893
 9999:  9999     1 42.23665
10000: 10000     1 57.38959

$timevar_data
$timevar_data$L
         id   date
      <int>  <num>
   1:     1   0.00
   2:     3   0.00
   3:     5   0.00
   4:     8   0.00
   5:    14   0.00
  ---
5119:  9988 421.80
5120:  9992 355.43
5121:  9995 246.33
5122:  9998 377.86
5123:  9998 483.76

$timevar_data$A
         id  date
      <int> <num>
   1:     1     0
   2:     3     0
   3:     5     0
   4:     7     0
   5:     9     0
  ---
5249:  9590   365
5250:  9671   365
5251:  9694   365
5252:  9859   365
5253:  9984   730
```

```
$outcome_data
         id    date
      <int>   <num>
   1:     1 1076.89
   2:     3   50.49
   3:     6  140.84
   4:     7  402.61
   5:    11  481.58
  ---
3341:  9979   60.41
3342:  9990   31.30
3343:  9993   19.82
3344:  9998  609.26
3345:  9999    8.99

$competing_data
         id   date
      <int>  <num>
   1:     4  89.71
   2:     8  91.81
   3:    10 545.28
   4:    13 446.94
   5:    14 459.49
  ---
3360:  9987  30.28
3361:  9988 476.90
3362:  9994 311.22
3363:  9996  71.81
3364:  9997 179.60

$censored_data
         id    date
      <int>   <num>
   1:     2  260.64
   2:     5    6.32
   3:     9  209.74
   4:    25  298.93
   5:    33  341.00
  ---
3287:  9989   53.65
3288:  9991  392.38
3289:  9992 1046.23
3290:  9995  758.63
3291: 10000  140.10
```

3 Example

3.1 Initializing

```
x <- rtmle_init(intervals = 3,name_id = "id",name_outcome = "Y",name_
    competing = "Dead",name_censoring = "Censored",censored_label = "
    censored")
x
```

Targeted minimum loss analysis of register data. Initialized to 3 time intervals (starting at time zero). The name of the subject id variable is 'id'. The outcome, competing risk and censoring variables are named 'Y', 'Dead', and 'Censored', respectively. TODO: The object contains no data yet. Add them with the functions 'add$_{\text{longdata}}$<-' and 'add$_{\text{data}}$<-'. TODO: The object contains no protocols. Add them with the function 'protocol<-'. TODO: The object contains no targets yet. Add them with the function 'target<-'.

3.2 Adding the data

```
x$long_data <- ld[c("outcome_data","censored_data","competing_data","
    timevar_data")]
baseline_data(x) <- ld$baseline_data[,start_followup_date:=0]
x
```

Targeted minimum loss analysis of register data. Initialized to 3 time intervals (starting at time zero). The name of the subject id variable is 'id'. The outcome, competing risk and censoring variables are named 'Y', 'Dead', and 'Censored', respectively. TODO: The object contains no protocols. Add them with the function 'protocol<-'. TODO: The object contains no targets yet. Add them with the function 'target<-'.

3.3 Defining the protocols

```
protocol(x) <- list(name = "always A",treatment_variables = "A",
    intervention = 1)
protocol(x) <- list(name = "never A",treatment_variables = "A",
    intervention = 0)
x
```

Targeted minimum loss analysis of register data. Initialized to 3 time intervals (starting at time zero). The name of the subject id variable is 'id'. The outcome, competing risk and censoring variables are named 'Y', 'Dead', and 'Censored', respectively. TODO: The object contains no targets yet. Add them with the function 'target<-'.

3.4 Defining the targets

```
target(x) <- list(name = "Outcome_risk",strategy = "additive",
    estimator = "tmle",time_horizon = 3,protocols = c("always A","
    never A"))
target(x) <- list(name = "Outcome_risk_g",strategy = "additive",
    estimator = "g",time_horizon = 3,protocols = c("always A","never
    A"))
x
```

Targeted minimum loss analysis of register data. Initialized to 3 time intervals (starting at time zero). The name of the subject id variable is 'id'. The outcome, competing risk and censoring variables are named 'Y', 'Dead', and 'Censored', respectively.

3.5 Sequential regression with and without TMLE update step

```
x <- long_to_wide(x,intervals = seq(0,2000,30.45*6))
prepare_data(x) <- list()
x <- run_rtmle(x,learn = "learn_glm")
```