Propensity of Treatment, Censoring and Competing risks

Alessandra Meddis
Section of Biostatistics

TMLE for Breakfast
May 23 2023

Censoring and competing event

So far, we considered the scenario without right censoring and competing event, this is not realistic in the real-world for time-to-event outcome.

- Right censoring is linked to loss to follow-up/drop out
  In an hypothetical target trial we would like to have no loss to follow-up
  $\rightarrow$ Causal question: Treatment effect if all patients had the specified treatment strategy without loss of follow-up.
- Competing event: event that prevents the event of interest to occur, for example death without cardiovascular disease (diabetes example)
  We cannot consider an hypothetical trial where we prevent people to die. This would never happen in reality.

Data structure

Right-censored data and competing event:
We assume the order collection : $(L(k-1), A(k-1), Y(k), D(k), C(k))$:

$$\mathcal{F}(Y(k)) = (\overline{Y}(k-1), \overline{D}(k-1), \overline{C}(k-1), \overline{A}(k-1), \overline{L}(k-1))$$
$$\mathcal{F}(C(k)) = (\overline{Y}(k), \overline{D}(k), \overline{C}(k-1), \overline{A}(k-1), \overline{L}(k-1))$$

We can factorize the data distribution in presence of censoring and competing event as follows:

$$P_X(x) = \prod_{k=1}^{K} P_{Y(k)|\mathcal{F}_{Y(k)}} P_{D(k)|\mathcal{F}_{D(k)}} P_{C(k)|\mathcal{F}_{C(k)}} \times P_{A(k-1)|\mathcal{F}_{A(k-1)}} P_{L(k-1)|\mathcal{F}_{L(k-1)}}$$

- Intervene on the censoring: estimate treatment effect **without** loss of follow-up: $G_{C(k)} := P_{C(k)|\mathcal{F}_{C(k)}}$
- $Q_{D(k)} := P_{D(k)|\mathcal{F}_{D(k)}}$

Average Treatment Effect

Target Parameter: $ATE(K) = \tilde{P}(Y^1(K) = 1) - \tilde{P}(Y^1(K) = 0)$
Under the idenitifiability assumptions for right-censored data we can write:
IPW identification:

$$\tilde{P}(Y^{a^*}(K) = 1) = P_X\left((Y(K)\left(\prod_{l=0}^{K-1} \frac{I(A(l) = a^*(l))\ I(C(l) = 1)}{G_{A(l)}(a^*(l))G_{C(l)}(1)}\right)\right)$$

- propensity of treatment: $G_{A(k)}(a^*(k))$ for the treatment strategy of interest
- propensity of of being uncensored: $G_{A(k)}(1)$

Estimation with right-censoring and competing event

- Nuisance parameters:
  - $G_{C(k)}(1)$, $G_{A(k)}(a^*)$ time-dependent propensity of treatment and censoring used as weights
  - $Q_{Y(k)}$ outcome model for the event of interest, where

$$D(k) = 1 \rightarrow Q_{Y(s)} = 0 \ \forall s \geq k$$

    by definition of competing event

In practice, K=2 with censoring and competing event

Consider 10 individuals with censoring C_k and competing event D_k. We are interested in the absilute risk for the treatment strategy $a^*(k) = 1 \ \forall k$

	A_0	C_1	Y_1	D_1	A_1	C_2	Y_2	D_2
1:	1	1	0	0	1	1	1	0
2:	0	1	1	0	0	1	1	0
3:	0	0	0	0	0	0	0	0
4:	1	0	0	0	1	0	0	0
5:	0	1	1	0	0	1	1	0
6:	0	1	0	0	0	1	0	0
7:	1	1	0	0	0	1	1	0
8:	1	1	0	0	1	0	0	0
9:	0	1	0	1	0	1	0	1
10:	0	1	0	0	0	1	0	0

In practice, K=2 with censoring and competing event

Consider 10 individuals with censoring C_k and competing event D_k. We are interested in the absilute risk for the treatment strategy $a^*(k) = 1 \ \forall k$

	A_0	C_1	Y_1	D_1	A_1	C_2	Y_2	D_2
1:	1	1	0	0	1	1	1	0
2:	0	1	1	0	0	1	1	0
3:	0	0	0	0	0	0	0	0
4:	1	0	0	0	1	0	0	0
5:	0	1	1	0	0	1	1	0
6:	0	1	0	0	0	1	0	0
7:	1	1	0	0	0	1	1	0
8:	1	1	0	0	1	0	0	0
9:	0	1	0	1	0	1	0	1
10:	0	1	0	0	0	1	0	0

In practice, K=2 with censoring and competing event

Consider 10 individuals with censoring C_k and competing event D_k. We are interested in the absolute risk for the treatment strategy $a^*(k) = 1 \ \forall k$

	A_0	C_1	Y_1	D_1	A_1	C_2	Y_2	D_2	
1:	1	1	0	0	1	1	1	0	✓
2:	0	1	1	0	0	1	1	0	
3:	0	0	0	0	0	0	0	0	
4:	1	0	0	0	1	0	0	0	
5:	0	1	1	0	0	1	1	0	
6:	0	1	0	0	0	1	0	0	
7:	1	1	0	0	0	1	1	0	
8:	1	1	0	0	1	0	0	0	✗
9:	0	1	0	1	0	1	0	1	
10:	0	1	0	0	0	1	0	0	

In practice, K=2 with censoring and competing event

	A_0	C_1	Y_1	D_1
1:	1	1	0	0
2:	0	1	1	0
3:	0	0	0	0
4:	1	0	0	0
5:	0	1	1	0
6:	0	1	0	0
7:	1	1	0	0
8:	1	1	0	0
9:	0	1	0	1
10:	0	1	0	0

1. $G_{A(0)}(1)* = \frac{I(A(0)=1)}{10} = 4/10 = 0.4$
2. $G_{C(1)}(1) = \frac{I(C(1)=1)}{10} = 8/10 = 0.8$

$$\rightarrow G_{A(0)}(1) \times G_{C(1)}(1) = 0.3$$

In practice, K=2 with censoring and competing event

	A_0	C_1	A_1	C_2
1:	1	1	1	1
6:	0	1	0	1
7:	1	1	0	1
8:	1	1	1	0
10:	0	1	0	1

When $Y(1) = 0, D(1) = 0$ (still at risk)

1. $G_{A(1)}(1) = \frac{I(\overline{A}(1)=1)}{I(\overline{A}(0)=1)} = 2/3 = 0.7$
2. $G_{C(2)}(1) = \frac{I(\overline{C}(2)=1)}{I(\overline{A}(1)=1, \overline{C}(0)=1)} = 1/2 = 0.5$

$\rightarrow G_{A(0)}(1) \times G_{C(1)}(1) G_{A(1)}(1) \times G_{C(2)}(1) = 0.1$

In practice, K=2 with censoring and competing event

When calling `ltmle`

```
> mod.ltmle$cum.g[,4]
      [,1] [,2] [,3] [,4]
 [1,]  0.4  0.3  0.2  0.1 #event at k=2, A0=A1=C0=C1=1
 [2,]  0.4  0.3  0.3  0.3 #event at k=1,  A0=0
 [3,]  0.4  0.3  0.2  0.1 #no event
 [4,]  0.4  0.3  0.2  0.1 #no event
 [5,]  0.4  0.3  0.3  0.3 #event at k=1, A0=0
 [6,]  0.4  0.3  0.2  0.1 #no event
 [7,]  0.4  0.3  0.2  0.1 #event at k=2, A1=0
 [8,]  0.4  0.3  0.2  0.1 #no event
 [9,]  0.4  0.3  0.3  0.3 #comp event k=1
[10,]  0.4  0.3  0.2  0.1 #no event
```

$\rightarrow \tilde{P}(Y^1(2)=1) = \frac{1}{10}\sum_{i=1}^{10}\frac{Y_i(2)}{w_i(2)} = \frac{1}{10}(\frac{1}{0.1}+0+0+...+0) = 1$

Specification in the `ltmle` package

To handle censoring:

- `Cnodes` : column names or indices in data of censoring nodes. Note that *0* means censored and *1* is uncensored (*censored*/ *uncensored* are also allowed)
- `survivalOutcome=TRUE`, so that the `Ynodes` are indicator of the event of interest. By default, if $Y(k) = 1 \rightarrow Y(s) = 1 \ \forall s \geq k$
- `gform`: list of formulas with same length of Anodes and Cnodes. Formulas specify the variables to be included in the model. If not specified, the entire history is considered.

To handle competing events:

- Introduce the competing event $D(k)$ in the `Lnodes` (time-dependent covariates)
- Specify the `deterministic.Q function` so that $D(k) = 1 \rightarrow Q_{Y(s)} = 0 \ \forall s \geq k$
- `gform` and Qform need to be specified so to exclude the competing event D on the list of covariates for the history

Call `ltmle` function

```
ltmle(data,
  Anodes=Anodes,
  Cnodes=Cnodes,
  Lnodes=Lnodes,
  Ynodes=Ynodes,
  abar=list(rep(1,10),rep(0,10)),
  gform=g.form,
  Qform=Q.form,
  deterministic.Q.function = det.Q.function,
  survivalOutcome = TRUE)
```

Comments

- we do not need any specification for the competing event, apart that $Y(K) = 0$ for individuals that occurred the competing event $\rightarrow$ exclude individuals from at risk set
- BE AWARE: `ltmle` consider bounds for the propensity values. By default `gbounds=(0.01,1)`
- BE AWARE: the order of columns in the input `data.frame` matters
  From the `ltmle` vignette:
  If there is a "block" of L and Y nodes not separated by A or C nodes, only one regression is required at the first L/Y node in a block.
  If we consider the order:

$$[L_0, A_0, C_1, D_1, Y_1, L_1, A_1, ..., C_K, Y_K]$$

  then: `Q.k.plus` model refers to D_k cause it is the first L/Y node of the block

$$\rightarrow [L_0, A_0, C_1, Y_1, L_1, D_1, A_1, ..., C_K, Y_K]$$

Definition of treatment strategy

Definition of the target parameter includes the definition of the treatment regimens of interest.

- Static treatment regimen: fixed treatment that is same for all individuals in the population, for example "always treated" and "never treated" strategies: $A(k) = a \ \forall k, \ a = 0, 1$
- Dynamic treatment regimen: treatment assignment depends on the individual's observed past, for example "start the treatment at a specific threshold of a biomarker" $L(k) \geq c \rightarrow A(s) = 1 \ \forall s > k$
- Stochastic treatment regimen: the treatment assignment is randomly generated accordingly to a user-specified probability distribution (at the moment not implemented in `ltmle`)

Definition of treatment in the `ltmle` package

- `Anodes`: column names or indices in data of treatment nodes
- `gform`: list of formulas with same length of Anodes. Formulas specified the variables to be included in the model. As default value, all parent nodes of L and A will be used.
- `SL.library`: list with one or more machine learning method. If specified `SuperLearner` will be called using the design matrix of `gform`. If NULL, `glm` will be called instead.
- `SL.cvControl`: number of folds for cross-validation, default value is 10
- `abar`: list of vectors for static treatment regimens of interest
- `rule`: function used to specify dynamic treatment regimens of interest

Call `ltmle` function

```
ltmle(data,
  Anodes=Anodes,
  Lnodes=Lnodes,
  Ynodes=Ynodes,
  abar=list(rep(1,10),rep(0,10))
  gform=NULL,
  SL.library="default",
  survivalOutcome = TRUE)
```