contTimeCausal - Continuous Time Causal Models

Shu Yang and Shannon T. Holloway

June 8, 2021

Introduction

contTimeCausal is an **R** package for evaluating the effect of time-varying treatments on a survival outcome. Currently, the package includes a structural failure time model (SFTM) and a Cox marginal structural model (MSM) that respect the continuous time nature of the underlying data processes in practice (i.e., the variables and processes are measured at irregularly spaced time points, which are not necessarily the same for all subjects).

The SFTM model assumes that the potential failure time, U, had the individual never received treatment and the observed failure time, T, follow

$$U \sim \int_0^T e^{\psi \times A_u} du,$$

where $\sim$ means "has the same distribution as" and A_u is the treatment indicator at time u.

The Cox MSM model assumes that the potential failure time, $T^{\overline{a}}$, under the treatment regime $\overline{a}$ (a hypothetical treatment regime from baseline to the event time) follows a proportional hazards model with the hazard function at u as

$$\lambda_0(u)e^{\psi \times a_u}.$$

It is assumed that individuals continuously received treatment until time V. The observed failure time can be censored. We assume an ignorable censoring mechanism in the sense that the censoring time is independent of the failure time given the treatment and covariate history.

The implemented methods can accommodate an arbitrary number of baseline and/or time-dependent covariates. If only time-independent covariates are included, the data required for analysis must contain the following information:

- **id**: A unique participant identifier.

- **U**: The time to the clinical event or censoring.

- **deltaU**: The clinical event indicator (1 if U is the event time; 0 otherwise).

- **V**: The time to optional treatment discontinuation, a clinical event, censoring, or a treatment-terminating event.

- **deltaV**: The indicator of optional treatment discontinuation (1 if treatment discontinuation was optional; 0 if treatment discontinuation was due to a clinical event, censoring or a treatment-terminating event).

If time-dependent covariates are to be included, the data must be a time-dependent dataset as described by package **survival**. Specifically, the time-dependent data must be specified for an interval (lower,upper] and the data must include the following additional information:

- **start**: The lower boundary of the time interval to which the data pertain.

- **stop**: The upper boundary of the time interval to which the data pertain.

The SFTM and CoxMSM are implemented through functions *ctSFTM()* and *ctCoxMSM()*, respectively. Both functions estimate the effect of the treatment regime (in terms of time to treatment discontinuation) for a survival outcome with time-varying treatment and confounding in the presence of dependent censoring. Function *ctSFTM()* estimates the effect under a continuous time SFTM, and *ctCoxMSM* estimates the effect under a Cox proportional hazards model. The functions return the estimated model parameter, ψ.

Functions

ctSFTM()

This function estimates the effect of time-varying treatments on a survival outcome under the SFTM. The value object returned is a continuous-time g-estimator of ψ.

The function call takes the following form:

```
ctSFTM(data, base = NULL, td = NULL)
```

where

- `data` is a data.frame object containing all required data as previously described. If only time-independent covariates are included in the model, the data.frame must use the following column headers: 'id', 'U', 'deltaU', 'V', and 'deltaV', which are defined above. If time-dependent covariates are to be included, the data.frame must be a time-dependent dataset as described by package **survival**. Specifically, the time-dependent data must be specified for an interval (start,stop] and the data must include two additional columns headers: 'start' and 'stop'.

- `base` is a character or numeric vector object or `NULL`. If baseline covariates are included in the model, `base` must be a vector containing the column headers or the column numbers of `data` that pertain to the time-independent covariates. If `NULL`, time-independent covariates are excluded from the model.

- `td` is a character or numeric vector object or `NULL` If time-dependent covariates are included in the model, `td` must be a vector containing the column headers or the column numbers of data that pertain to the time-dependent covariates. If `NULL`, time-dependent covariates are excluded from the model. Note that this vector should <u>not</u> include required columns 'start' and 'stop'.

Note that only one of `base` and `td` can be specified as `NULL`.

The value object returned by *ctSFTM()* an S3 object of class **ctc**, containing $psi, the estimated model parameter, and $coxPH, the Cox regression for V.

ctCoxMSM()

This function estimates the effect of time-varying treatments on a survival outcome under the Cox proportional hazards model. The value object returned is an inverse probability of treatment weighting estimator of ψ.

The function call takes the following form:

```
ctCoxMSM(data, base = NULL, td = NULL)
```

where the inputs are as defined above for *ctSFTM()* and their descriptions are not repeated here.

The value object returned by *ctCoxMSM()* is also an S3 object of class **ctc**, containing $psi, the estimated model parameter, and $coxPH, the Cox regression for V.

print()

A convenience function for displaying the primary results of the analysis.

The function call takes the following form:

```
print(x, ...)
```

where x is the **ctc** object returned by *ctSFTM()* or *ctCoxMSM()*.

Examples

To illustrate the call structure and results of *ctSFTM()* and *ctCoxMSM()*, we use the dataset provided with the package, `ctcData`. This dataset was generated only for the purposes of illustrating the package and should not be interpreted as representing a real-world dataset. The dataset contains the following observations for 1,000 participants:

- id A unique participant identifier.

- start The left side of the time interval for time-dependent covariate xt.

- stop The right side of the time interval for time-dependent covariate xt.

- xt A continuous time-dependent covariate.

- x A continuous baseline covariate.

- deltaU A binary indicator of the clinical event. If the clinical event occurred, takes value 1; otherwise 0.

- deltaV A binary indicator of treatment discontinuation. If treatment discontinuation was optional, takes value 1. If treatment discontinuation was due to the clinical event, censoring, or a treatment-terminating event, takes value 0.

- U The time to the clinical event or censoring.

- V The time to optimal treatment discontinuation, the clinical event, censoring, or a treatment-terminating event.

The data can be loaded in the usual way

```
data(ctcData)
```

```
head(ctcData)
```

```
##   id start stop        xt x deltaU deltaV     U    V
## 1  1     0 1.41 -5.093096 1      1      0  1.41 1.41
## 2  2     0 2.00 -1.660851 1      1      0  2.00 2.00
## 3  3     0 2.95 -4.500386 1      1      1  2.95 2.90
## 4  4     0 0.22 -4.170195 1      1      0  0.22 0.22
## 5  5     0 2.99 -3.039027 1      1      0  2.99 2.99
## 6  6     0 5.00 -2.173769 0      1      1 14.81 5.52
```

Before considering the summary statistics of the dataset, we break the set into time-independent and time-dependent components. The time-independent data can be extracted as follows

```
ti <- ctcData[,c(5L:9L)]
ti <- ti %>% distinct()
```

where we have eliminated duplicate rows using **dplyr**'s *distinct()* function. The time-dependent component is trivially extracted

```
td <- ctcData[,2L:4L]
```

The summary statistics for the time-independent data

```
summary(object = ti)
```

```
##        x            deltaU           deltaV              U
##  Min.   :0.000   Min.   :0.000   Min.   :0.0000   Min.   : 0.010
##  1st Qu.:0.000   1st Qu.:1.000   1st Qu.:0.0000   1st Qu.: 1.660
##  Median :1.000   Median :1.000   Median :0.0000   Median : 3.470
##  Mean   :0.552   Mean   :0.856   Mean   :0.3463   Mean   : 4.692
##  3rd Qu.:1.000   3rd Qu.:1.000   3rd Qu.:1.0000   3rd Qu.: 6.430
##  Max.   :1.000   Max.   :1.000   Max.   :1.0000   Max.   :28.950
##        V
##  Min.   : 0.010
##  1st Qu.: 1.070
##  Median : 2.360
##  Mean   : 3.119
##  3rd Qu.: 4.530
##  Max.   :14.490
```

show that the binary baseline covariate x is approximately evenly distributed across the participants; that
$\sim 87\%$ of the participants experienced a clinical event; and that for $\sim 30\%$ of the participants treatment was
optionally discontinued. The maximum time to a clinical event or censoring is 28.95. Considering only those
participants that optionally discontinued treatment

```
summary(object = ti$V[ti$deltaV==1L])
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.010   0.920   2.060   2.815   3.840  13.670
```

we see that the $0.01 \le V \le 13.67$.

From the summary of the time-dependent data, we see that xt is a continuous variable in the range
$-7.027 \le xt \le 2.736$. Measurements were taken at time points $\{0, 5, 10\}$.

```
summary(object = td)
```

```
##      start           stop              xt
##  Min.   : 0.00   Min.   : 0.010   Min.   :-7.027
##  1st Qu.: 0.00   1st Qu.: 2.090   1st Qu.:-3.673
##  Median : 0.00   Median : 5.000   Median :-2.741
##  Mean   : 1.77   Mean   : 4.858   Mean   :-2.667
##  3rd Qu.: 5.00   3rd Qu.: 5.883   3rd Qu.:-1.785
##  Max.   :10.00   Max.   :28.950   Max.   : 2.736
```

In the first example, we estimate the parameter of the continuous time SFTM using all of the available data
as follows

```
res <- ctSFTM(data = ctcData, base = "x", td = "xt")
```

```
## Estimated psi: 1.0797
```

The value object returned is an S3 object containing the estimated parameter and the Cox regression for
V. As a precaution, we encourage users to verify that the parameters included in the Cox regression are as
expected from the input. Here, we see that both x and xt were included.

```
res
```

```
## form: Surv(start, stop, V.time) ~ x + xt

## Call:
## survival::coxph(formula = form, data = dataTD)
##
##      coef exp(coef) se(coef)     z        p
## x  0.19483   1.21511  0.11647 1.673 0.094352
## xt 0.14580   1.15697  0.04349 3.352 0.000801
##
## Likelihood ratio test=13.86  on 2 df, p=0.0009756
## n= 221888, number of events= 303
##
##
## Estimated psi:  1.0797
```

To include only the baseline covariates in the model, we do not change input `td` from its default setting of NULL.

```
ctSFTM(data = ctcData, base = "x")
```

```
## Estimated psi: 1.2726

## form: Surv(start, stop, V.time) ~ x

## Call:
## survival::coxph(formula = form, data = dataTD)
##
##     coef exp(coef) se(coef)    z      p
## x 0.1992    1.2204   0.1165 1.71 0.0873
##
## Likelihood ratio test=2.95   on 1 df, p=0.08611
## n= 221888, number of events= 303
##
##
## Estimated psi:  1.2726
```

Note that only x was included in the Cox regression.

Similarly, to include only the time-dependent covariates, we do not provide input `base`.

```
ctSFTM(data = ctcData, td = "xt")
```

```
## Estimated psi: 1.0777

## form: Surv(start, stop, V.time) ~ xt

## Call:
## survival::coxph(formula = form, data = dataTD)
##
##      coef exp(coef) se(coef)     z        p
## xt 0.14655   1.15784  0.04347 3.371 0.000748
##
## Likelihood ratio test=11.04  on 1 df, p=0.0008893
## n= 221888, number of events= 303
##
##
```

```
## Estimated psi:  1.0777
```

Next we estimate ψ under the continuous time Cox MSM using all of the available data

```
res <- ctCoxMSM(data = ctcData, base = "x", td = "xt")
```

```
## Estimated psi: 1.0497
```

The value object returned is an S3 object containing the estimated parameter and the Cox regression for V. Notice that the regression results are the same as those obtained above for the full SFTM analysis. The estimated parameter, ψ, is similar to that obtained under the SFTM with both time-dependent and time-independent covariates.

```
res
```

```
## form: Surv(start, stop, V.time) ~ x + xt

## Call:
## survival::coxph(formula = form, data = dataTD)
##
##        coef exp(coef) se(coef)     z        p
## x   0.19483   1.21511  0.11647 1.673 0.094352
## xt  0.14580   1.15697  0.04349 3.352 0.000801
##
## Likelihood ratio test=13.86  on 2 df, p=0.0009756
## n= 221888, number of events= 303
##
##
## Estimated psi:  1.0497
```

As for the `ctSFTM()`, we can also consider only the time-independent covariates

```
ctCoxMSM(data = ctcData, base = "x")
```

```
## Estimated psi: 1.2431
## form: Surv(start, stop, V.time) ~ x

## Call:
## survival::coxph(formula = form, data = dataTD)
##
##      coef exp(coef) se(coef)    z      p
## x  0.1992    1.2204   0.1165 1.71 0.0873
##
## Likelihood ratio test=2.95  on 1 df, p=0.08611
## n= 221888, number of events= 303
##
##
## Estimated psi:  1.2431
```

Or, to include only the time-dependent covariates

```
ctCoxMSM(data = ctcData, td = "xt")
```

```
## Estimated psi: 1.042
```

```
## form: Surv(start, stop, V.time) ~ xt
## Call:
## survival::coxph(formula = form, data = dataTD)
##
##        coef exp(coef) se(coef)     z        p
## xt 0.14655   1.15784   0.04347 3.371 0.000748
##
## Likelihood ratio test=11.04  on 1 df, p=0.0008893
## n= 221888, number of events= 303
##
##
## Estimated psi:  1.042
```

References

Yang, S., A. A. Tsiatis, and M. Blazing (2018). Modeling survival distribution as a function of time to treatment discontinuation: A dynamic treatment regime approach, *Biometrics*, 74, 900–909.

Yang, S., K. Pieper, and F. Cools (2020). Semiparametric estimation of structural failure time model in continuous-time processes. *Biometrika*, 107, 123–136.