ConCR-TMLE R Paper

David Chen, Thomas Gerds, Helene Rytgaard

Abstract

An abstract of less than 150 words.

Introduction

Data Structure

Consider a survival analysis on an interval $[0, t_{max}]$ with competing risks. Let T_j^a denote counterfactual time-to-event variables for event j and intervention a, for competing events $j \in \mathcal{J} = \{1, 2, \ldots, J\}$ and an intervention $a \in \mathcal{A}$. Our counterfactual data structure can then be denoted by

$$(T_j^a,\ X\ :\ a \in \mathcal{A},\ j \in \mathcal{J})$$

where $X \in \mathbb{R}^d$ is a d-dimensional vector of baseline covariates. For a single time-point binary intervention, as in many randomized control trials, $\mathcal{A} = \{0, 1\}$ and the corresponding counterfactual data is

$$(T_j^1,\ T_j^0,\ X\ :\ j \in \mathcal{J})$$

```
head(counterfactuals)
```

	T.j1.a0	T.j1.a1	T.j2.a0	T.j2.a1	L1	L2	L3
1	0.1599887	0.4906215	0.5399409	0.5803671	-1.7677221	4	3.0093952
2	1.1369533	1.9210028	0.2375033	0.9133089	-0.4916921	0	0.3294865
3	0.3447736	1.2538906	0.4779721	0.8540658	0.3214659	3	4.1630246
4	4.6631762	0.3718961	1.5650534	0.2485393	1.4606608	3	1.5313713
5	0.1430018	0.5951058	0.3003895	0.9765322	1.5372426	2	1.5580743
6	1.8419819	3.9131870	1.8517334	3.0117075	-0.3395685	4	0.8455748

Let O denote the corresponding coarsened observed data where $O \sim P_0$. The observed data would include the time-to-censoring C, and observed intervention A. The time to first event (censoring or otherwise) we denote as $\widetilde{T} = \min(C,\ T_j\ :\ j \in \mathcal{J})$ with $\Delta = (\operatorname*{argmin}_j T_j) \times \mathbf{1}(\min_j T_j \leq C)$ marking which outcome is observed ($\Delta = 0$ being that censoring occurred). The observable right-censored survival data with competing events can then be represented as

$$O = (\widetilde{T},\ \Delta,\ A,\ X)$$

```
head(observed)
```

	T.tilde	Delta	A	L1	L2	L3
1	0.20711055	0	1	-1.7677221	4	3.0093952
2	0.91147298	0	1	-0.4916921	0	0.3294865
3	0.08374201	0	0	0.3214659	3	4.1630246
4	0.29772679	0	0	1.4606608	3	1.5313713
5	0.14300179	1	0	1.5372426	2	1.5580743
6	1.06839386	0	0	-0.3395685	4	0.8455748

This observed data also allows the "long-format" formulation, where with single time-point intervention variable A and baseline covariate vector X,

$$O = (N_j(t),\ N_c(t),\ A,\ X\ :\ j \in \mathcal{J}, t \leq \widetilde{T})$$

Here $N_j(t) = \mathbf{1}(\widetilde{T} \leq t, \Delta = j)$ and $N_c(t) = \mathbf{1}(\widetilde{T} \leq t, \Delta = 0)$ denote counting processes for event j and censoring respectively.

Under coarsening at random (CAR), the observed data likelihood can be factorized as

$$\begin{aligned} p(O) = p(X)\, \pi(A \mid X)\, \lambda_c(\widetilde{T} \mid A,\, X)^{\mathbf{1}(\Delta=0)} S_c(\widetilde{T}\text{-} \mid A,\, X) \\ \prod_{j=1}^{J} S(\widetilde{T}\text{-} \mid A,\, X)\, \lambda_j(\widetilde{T} \mid A,\, X)^{\mathbf{1}(\Delta=j)} \end{aligned}$$

where $\lambda_c(t \mid A,\, X)$ is the hazard of the censoring process and $\lambda_j(t \mid A,\, X)$ is the hazard of the j^{th} event process. Additionally

$$S_c(t \mid a,\, x) = \exp\left(-\int_0^t \lambda_c(s \mid a,\, x)\, ds\right)$$

while in a pure competing risks setting

$$S(t \mid a,\, x) = \exp\left(-\int_0^t \sum_{j=1}^{J} \lambda_j(s \mid a,\, x)\, ds\right)$$

and

$$\begin{aligned} F_j(t \mid a,\, x) &= \int_0^t S(s\text{-} \mid a,\, x) \lambda_j(s \mid a,\, x)\, ds \\ &= \int_0^t \exp\left(-\int_0^s \sum_{j=1}^{J} \lambda_j(u \mid a,\, x)\, du\right) \lambda_j(s \mid a,\, x)\, ds. \end{aligned}$$

Target Parameter

Given the identification assumptions of

1. Consistency : $T = T^a$ when $A = a$ for $a = 0, 1$.
2. No unmeasured confounding: $T^a \perp\!\!\!\perp A \mid X$ for $a = 0, 1$.
3. Coarsening at random on censoring: $T \perp\!\!\!\perp C \mid A,\, X$

the hypothetical distribution for data generated following a desired treatment regime involving $A \sim \pi^*(A \mid X)$ and the prevention of the censoring process can be identified as

$$p^{\pi^*}(O) = p(X)\, \pi^*(A \mid X) \prod_{j=1}^{J} S(\widetilde{T}\text{-} \mid A,\, X) \lambda_j(\widetilde{T} \mid A,\, X)^{\mathbf{1}(\Delta=j)}$$

For a target parameter of the cause $k \in J$ absolute risk at time τ under this treatment regime π^*, the corresponding efficient influence function $D^*_{\pi^*,k,\tau}(P)(O)$ is

$$\sum_{j=1}^{J} \int_0^\tau \left[\frac{\pi^*(A \mid X)\,\mathbf{1}(s \leq \tau)}{\pi(A \mid X)\, S_c(s\text{-} \mid A,\, X)} \left(\mathbf{1}(\delta = k) - \frac{F_k(\tau \mid A,\, X) - F_k(s \mid A,\, X)}{S(s \mid A,\, X)}\right)\right.$$
$$\left.\left(N_j(ds) - \mathbf{1}(\widetilde{T} \geq s)\, \lambda_j(s \mid A,\, X)\right)\right] ds$$
$$+ \sum_{a=0,1} F_k(t \mid A = a, X)\, \pi^*(a \mid X) - \Psi_{\pi^*,k,\tau}(P_0)$$

with a clever covariate $h_{\pi^*,k,j,\tau,s}$

$$h_{\pi^*,\,k,\,j,\,\tau}(s) = \frac{\pi^*(A \mid X)\,\mathbf{1}(s \leq \tau)}{\pi(A \mid X) S_c(s\text{-} \mid A,\, X)} \left(\mathbf{1}(\delta = k) - \frac{F_k(\tau \mid A,\, X) - F_k(s \mid A,\, X)}{S(s \mid A,\, X)}\right)$$

In the binary point treatment case, for the cause k absolute risk at time τ if all individuals had been assigned to the treatment condition, $\pi^* = (A = 1)$, we would have

$$D^*_{1,k,\tau}(P)(O) =$$
$$\sum_{j=1}^{J} \int_0^\tau \left[\frac{\mathbf{1}(A = 1)\,\mathbf{1}(s \leq \tau)}{\pi(A \mid X)\, S_c(s\text{-} \mid A,\, X)} \left(\mathbf{1}(\delta = k) - \frac{F_k(\tau \mid A,\, X) - F_k(s \mid A,\, X)}{S(s \mid A,\, X)}\right)\right.$$
$$\left.\left(N_j(ds) - \mathbf{1}(\widetilde{T} \geq s)\, \lambda_j(s \mid A,\, X)\right)\right] ds$$
$$+ F_k(t \mid A = 1,\, X) - \Psi_{\pi^*,k,\tau}(P_0)$$

with a clever covariate $h_{\pi^*,k,j,\tau,s}$

$$h_{1,\,k,\,j,\,\tau}(s) = \frac{\mathbf{1}(A = 1)\,\mathbf{1}(s \leq \tau)}{\pi(A \mid X) S_c(s\text{-} \mid A,\, X)} \left(\mathbf{1}(\delta = k) - \frac{F_k(\tau \mid A,\, X) - F_k(s \mid A,\, X)}{S(s \mid A,\, X)}\right)$$

For estimation of survival-curve derived estimands such as the cause-specific absolute risks, the components of the data distribution that must be estimated are $g(A \mid X)$ and $S_c(t \mid A,\, X)$, $\lambda_j(t \mid A,\, X)$, $F_j(t \mid A,\, X)$, and $S(t \mid A,\, X)$

Estimation

Cross-Validation Specification

Let $D_n = \{O_i\}_{i=1}^n$ be an observed sample of n i.i.d observations of $O \sim P_0$. For V-fold cross validation, let $B_n = \{1, ..., V\}^n$ be a random vector that assigns the n observations into V validation folds. For each $v \in \{1, ..., V\}$ we then define training set $D_v^{\mathcal{T}} = \{O_i : B_n(i) = v\}$ with the corresponding validation set $D_v^{\mathcal{V}} = \{O_i : B_n(i) \neq v\}$.

Stratified Cross-Validation

```
StrataIDs <- factor(paste(observed[["A"]], observed[["Delta"]]))
CVFolds <- origami::make_folds(n = observed,
                   fold_fun = origami::folds_vfold,
                   strata_ids = StrataIDs)
```

Propensity Score Estimation

For the true conditional distribution of A given X, $\pi_0(\cdot \mid X)$, and $\hat{\pi} : D_n \to \hat{\pi}(D_n)$, let L_π be a loss function such that the risk $\mathbb{E}_0 [L_\pi(\hat{\pi}, O)]$ is minimized when $\hat{\pi} = \pi_0$. For instance, with a binary A, we may specify the negative log loss $L_\pi(\hat{\pi}, O) = -\log\left(\hat{\pi}(1 \mid X)^A \, \hat{\pi}(0 \mid X))^{1-A}\right)$. We can then define the discrete superlearner selector which chooses from a set of candidate models $\mathcal{M}_\pi$ the candidate propensity score model that has minimal cross validated risk

$$\hat{\pi}^{SL} = \operatorname*{argmin}_{\hat{\pi} \in \mathcal{M}_\pi} \sum_{v=1}^{V} P_{D_v^{\mathcal{V}}} \; L_\pi(\hat{\pi}(D_v^{\mathcal{T}}), D_v^{\mathcal{V}})$$

This discrete superlearner model $\hat{\pi}^{SL}$ is then fitted on the full observed data D_n and used to estimate $\pi_0(A \mid X)$

```
CovDataTable <- observed[, -c("T.tilde", "Delta", "A")]
Models <- list("Trt" = sl3::make_learner(sl3:::Lrnr_glm))
Intervention <- list(
  "A=1" = list("intervention" = function(a, L) rep_len(1, length(a)),
           "g.star" = function(a, L) {as.numeric(a == 1)}),
  "A=0" = list("intervention" = function(a, L) rep_len(0, length(a)),
           "g.star" = function(a, L) {as.numeric(a == 0)})
)

RegsOfInterest <- getRegsOfInterest(Intervention = Intervention,
                    Treatment = observed[["A"]],
                    CovDataTable = CovDataTable)

PropScores <- getPropScore(Treatment = observed[["A"]],
               CovDataTable = CovDataTable,
               Models = Models,
               MinNuisance = 0.05,
               RegsOfInterest = RegsOfInterest,
               PropScoreBackend = "sl3",
               CVFolds = CVFolds)
```

Hazard Estimation

Let $\lambda_{0,\,\delta}$ be the true censoring and cause-specific hazards when $\delta = 0$ and $\delta = 1, \ldots, J$ respectively. Let $\mathcal{M}_\delta$ for $\delta = 0, \ldots, J$ be the sets of candidate models,

$\{\hat{\lambda}_\delta : D_n \to \hat{\lambda}_\delta(D_n)\}$, for the censoring and cause-specific hazards and let L_δ be loss functions such that the risks $\mathbb{E}_0\left[L_\delta(\hat{\lambda}_\delta, O)\right]$ are minimized when $\hat{\lambda}_\delta = \lambda_{0,\delta}$, for instance log likelihood loss. We can then define the discrete superlearner selectors for each δ which choose from the set of candidate models $\mathcal{M}_\delta$ the candidate propensity score model that has minimal cross validated risk

$$\hat{\lambda}_\delta^{SL} = \underset{\hat{\lambda}_\delta \in \mathcal{M}_\delta}{\operatorname{argmin}} \sum_{v=1}^{V} P_{D_v^{\mathcal{V}}} \; L_\pi(\hat{\lambda}_\delta(D_v^{\mathcal{T}}), D_v^{\mathcal{V}})$$

These discrete superlearner selections $\hat{\lambda}_\delta^{SL}$ are then fitted on the full observed data D_n and used to estimate $\lambda_\delta(t \mid A,\, X)$, $F_\delta(t \mid A,\, X)$, $S(t \mid A,\, X)$, and $S_c(t\text{-} \mid A,\, X)$ for $j = 1, \dots, J$.

Lagged Censoring Survival

Let $\mathcal{S} = \{s_1, s_2, \dots, s_m\}$ be the set containing all target and observed event times, ordered such that $s_1 < s_2 < \cdots < s_m$. Then for all $s \in \mathcal{S}$ we compute

$$\hat{S}_c(s\text{-} \mid A,\, X) = \prod_{s_i < s} \left(1 - \hat{\lambda}_0^{SL}(s_i \mid A,\, X)\right)$$

Cause-Specific Hazards, Event-Free Survival, and Cause-Specific Absolute Risks

For $j = 1, \dots, J$ and $s \in \mathcal{S}$, the super learner selections $\hat{\lambda}_j^{SL}$ are fit on the full observed data D_n, and used to compute the event free survival

$$\hat{S}(s \mid A,\, X) = \exp\left(\text{-} \sum_{s_i \le s} \sum_{j=1}^{J} \hat{\lambda}_j^{SL}(s_i \mid A,\, X)\right)$$

cause-specific absolute risks

$$\hat{F}_j(s \mid A,\, X) = \sum_{s_i \le s} \hat{S}(s_i \mid A,\, X)\, \hat{\lambda}_j^{SL}(s_i \mid A,\, X)$$

Computing the Efficient Influence Function

For each desired treatment regime π^*, each target time τ, and each target event k, the efficient influence functions for each individual are computed in parts.

Clever Covariate: Nuisance Weight

For every $s_i \in \mathcal{S}$

$$NW_i = \frac{1}{\pi(a \mid x) S_c(s_i\text{-} \mid a,\, x)}$$

1 nuisance weight for every individual at every time $s_i \in \mathcal{S}$

Clever Covariate $h_{\pi^*,k,j,\tau}(s_i)$

The stored cause-specific hazards $\hat{\lambda}_j^{SL}(s_i \mid a,\, x)$ and event-free survival $\hat{S}(s_i \mid a,\, x)$ are used to calculate the cause-specific absolute risks $\hat{F}_j(s_i \mid a,\, x)$, then combined with the nuisance weight to calculate the clever covariates.

$$h_{\pi^*,\,k,\,j,\,\tau}(s_i) = \pi^*(a \mid x)\,\mathbf{1}(s_i \leq \tau) \times \mathsf{NW}_i \times \left(\mathbf{1}(\delta = k) - \frac{F_k(\tau \mid a,\, x) - F_k(s_i \mid a,\, x)}{S(s_i \mid a,\, x)}\right)$$

1 clever covariate for every individual, for every regime of interest, for every target event, for every target time, at every time $s_i \in \mathcal{S}$.

EIC

The sum over events and over time are done in a per person loop, the addition of the absolute risk and subtraction of the target estimand are done later, outside of the loop.

$$\sum_{j=1}^{J} \int_0^{\tau \wedge \tilde{t}} h_{\pi^*,\,k,\,j,\,\tau}(s) \times \left(N_j(ds) - \mathbf{1}(\widetilde{T} \geq s)\,\lambda_j(s \mid A,\, X)\right)\, ds$$

$$+F_k(t \mid A = \pi^*,\, X) - \Psi_{\pi^*,k,\tau}(P_0)$$