TMLE update step: the fluctuation model

TMLE for breakfast May 23

Mark Bech Knudsen
Section of Biostatistics

UNIVERSITY OF COPENHAGEN

Note

Maybe the most technical talk of the workshop.

Note

Maybe the most technical talk of the workshop.

- Explain *how* we update our initial estimates in TMLE.

Note

Maybe the most technical talk of the workshop.

- Explain *how* we update our initial estimates in TMLE.
- Give an idea of *why* we do it like this.

Note

Maybe the most technical talk of the workshop.

- Explain *how* we update our initial estimates in TMLE.
- Give an idea of *why* we do it like this.
- R example at the end.

Recap: Sequential regression

Want to estimate $\Psi_0 = \Psi(P_{\mathbf{X}}) \overset{!}{=} \tilde{P}(Y^{(\bar{a})} = 1)$, the probability of experiencing the outcome before or at time K if the subject receives treatment according to regime $\bar{a}$. Assuming the identifiability conditions to hold, we can estimate this by sequential regression:

Recap: Sequential regression

Want to estimate $\Psi_0 = \Psi(P_{\mathbf{X}}) \stackrel{!}{=} \tilde{P}(Y^{(\bar{a})} = 1)$, the probability of experiencing the outcome before or at time K if the subject receives treatment according to regime $\bar{a}$. Assuming the identifiability conditions to hold, we can estimate this by sequential regression:

1. Fit model using $Y = Y(K)$ as outcome, conditional on the past of $Y(K)$.
   $\rightarrow$ Predictions $\hat{\bar{Q}}^{\bar{a}}_{L(K)}$ according to treatment regime $\bar{a}$.

Recap: Sequential regression

Want to estimate $\Psi_0 = \Psi(P_{\mathbf{X}}) \overset{!}{=} \tilde{P}(Y^{(\bar{a})} = 1)$, the probability of experiencing the outcome before or at time K if the subject receives treatment according to regime $\bar{a}$. Assuming the identifiability conditions to hold, we can estimate this by sequential regression:

1. Fit model using $Y = Y(K)$ as outcome, conditional on the past of $Y(K)$.
   $\rightarrow$ Predictions $\hat{\bar{Q}}^{\bar{a}}_{L(K)}$ according to treatment regime $\bar{a}$.
2. Fit model using $\hat{\bar{Q}}^{\bar{a}}_{L(K)}$ as outcome, conditional on the past of $Y(K-1)$.
   $\rightarrow$ Predictions $\hat{\bar{Q}}^{\bar{a}}_{L(K-1)}$ according to treatment regime $\bar{a}$.

Recap: Sequential regression

Want to estimate $\Psi_0 = \Psi(P_{\mathbf{X}}) \stackrel{!}{=} \tilde{P}(Y^{(\bar{a})} = 1)$, the probability of experiencing the outcome before or at time K if the subject receives treatment according to regime $\bar{a}$. Assuming the identifiability conditions to hold, we can estimate this by sequential regression:

1. Fit model using $Y = Y(K)$ as outcome, conditional on the past of $Y(K)$.
   $\rightarrow$ Predictions $\hat{\bar{Q}}^{\bar{a}}_{L(K)}$ according to treatment regime $\bar{a}$.
2. Fit model using $\hat{\bar{Q}}^{\bar{a}}_{L(K)}$ as outcome, conditional on the past of $Y(K-1)$.
   $\rightarrow$ Predictions $\hat{\bar{Q}}^{\bar{a}}_{L(K-1)}$ according to treatment regime $\bar{a}$.
3. Repeat until predictions $\hat{\bar{Q}}^{\bar{a}}_{L(1)}$.

Recap: Sequential regression

Want to estimate $\Psi_0 = \Psi(P_{\mathbf{X}}) \stackrel{!}{=} \tilde{P}(Y^{(\bar{a})} = 1)$, the probability of experiencing the outcome before or at time K if the subject receives treatment according to regime $\bar{a}$. Assuming the identifiability conditions to hold, we can estimate this by sequential regression:

1. Fit model using $Y = Y(K)$ as outcome, conditional on the past of $Y(K)$.
   $\rightarrow$ Predictions $\hat{\bar{Q}}^{\bar{a}}_{L(K)}$ according to treatment regime $\bar{a}$.
2. Fit model using $\hat{\bar{Q}}^{\bar{a}}_{L(K)}$ as outcome, conditional on the past of $Y(K-1)$.
   $\rightarrow$ Predictions $\hat{\bar{Q}}^{\bar{a}}_{L(K-1)}$ according to treatment regime $\bar{a}$.
3. Repeat until predictions $\hat{\bar{Q}}^{\bar{a}}_{L(1)}$.
4. Compute $\hat{\Psi}^{\text{SR}} = \hat{\bar{Q}}^{\bar{a}}_{L(0)} = \frac{1}{n}\sum_{i=1}^{n} \hat{\bar{Q}}^{\bar{a}}_{L(1),i}$

What are we targeting?

In each step, we fit a model of outcome Y or predictions $\hat{\hat{Q}}^{\bar{a}}_{L(k)}$ conditional on the past.

What are we targeting?

In each step, we fit a model of outcome Y or predictions $\hat{\bar{Q}}^{\bar{a}}_{L(k)}$ conditional on the past.

These regressions provide the best possible fit for the parameters $\bar{Q}^{\bar{a}}_{L(k)}$ of our model according to some loss function.

What are we targeting?

In each step, we fit a model of outcome Y or predictions $\hat{\bar{Q}}^{\bar{a}}_{L(k)}$ conditional on the past.

These regressions provide the best possible fit for the parameters $\bar{Q}^{\bar{a}}_{L(k)}$ of our model according to some loss function.

Observation: Optimal estimation of $\bar{Q}^{\bar{a}}_{L(k)}$ does not necessarily imply optimal (efficient) estimation of the target parameter $\Psi_0 = \tilde{P}(Y^{(\bar{a})} = 1)$.

What are we targeting?

In each step, we fit a model of outcome Y or predictions $\hat{\bar{Q}}^{\bar{a}}_{L(k)}$ conditional on the past.

These regressions provide the best possible fit for the parameters $\bar{Q}^{\bar{a}}_{L(k)}$ of our model according to some loss function.

Observation: Optimal estimation of $\bar{Q}^{\bar{a}}_{L(k)}$ does not necessarily imply optimal (efficient) estimation of the target parameter $\Psi_0 = \tilde{P}(Y^{(\bar{a})} = 1)$.

Challenge: How do we then obtain an efficient estimator of Ψ_0?

Influence functions

If $G_{\bar{A}(k)}(\mathbf{X}) = \prod_{i=0}^{k} G_{A(i)}(\mathbf{X}), G_{\bar{C}(k)}(\mathbf{X}) = \prod_{i=1}^{k} G_{C(i)}(\mathbf{X})$, we know that

$$\Psi_0 = \Psi(P_{\mathbf{X}}) = \mathbb{E}_{P_{\mathbf{X}}} \left\{ Y(K) \frac{\mathbb{1}(\bar{A}(K-1) = \bar{a}(K-1), \bar{C}(K) = 1)}{G_{\bar{A}(K-1)}(\mathbf{X}) G_{\bar{C}(K)}(\mathbf{X})} \right\}$$

Influence functions

If $G_{\bar{A}(k)}(\mathbf{X}) = \prod_{i=0}^{k} G_{A(i)}(\mathbf{X})$, $G_{\bar{C}(k)}(\mathbf{X}) = \prod_{i=1}^{k} G_{C(i)}(\mathbf{X})$, we know that

$$\Psi_0 = \Psi(P_{\mathbf{X}}) = \mathbb{E}_{P_{\mathbf{X}}}\left\{ Y(K)\frac{\mathbb{1}(\bar{A}(K-1) = \bar{a}(K-1), \bar{C}(K) = 1)}{G_{\bar{A}(K-1)}(\mathbf{X})\, G_{\bar{C}(K)}(\mathbf{X})} \right\}$$

Replacing $\mathbb{E}_{P_{\mathbf{X}}}$ with the empirical average yields the estimator

$$\hat{\Psi}^{\mathrm{IPTW}} = \frac{1}{n}\sum_{i=1}^{n} Y_i(K)\frac{\mathbb{1}(\bar{A}_i(K-1) = \bar{a}(K-1), \bar{C}_i(K) = 1)}{G_{\bar{A}(K-1)}(\mathbf{X}_i)\, G_{\bar{C}(K)}(\mathbf{X}_i)}$$

Influence functions

If $G_{\bar{A}(k)}(\mathbf{X}) = \prod_{i=0}^{k} G_{A(i)}(\mathbf{X})$, $G_{\bar{C}(k)}(\mathbf{X}) = \prod_{i=1}^{k} G_{C(i)}(\mathbf{X})$, we know that

$$\Psi_0 = \Psi(P_{\mathbf{X}}) = \mathbb{E}_{P_{\mathbf{X}}} \left\{ Y(K) \frac{\mathbb{1}(\bar{A}(K-1) = \bar{a}(K-1), \bar{C}(K) = 1)}{G_{\bar{A}(K-1)}(\mathbf{X}) G_{\bar{C}(K)}(\mathbf{X})} \right\}$$

Replacing $\mathbb{E}_{P_{\mathbf{X}}}$ with the empirical average yields the estimator

$$\hat{\Psi}^{\mathsf{IPTW}} = \frac{1}{n} \sum_{i=1}^{n} Y_i(K) \frac{\mathbb{1}(\bar{A}_i(K-1) = \bar{a}(K-1), \bar{C}_i(K) = 1)}{G_{\bar{A}(K-1)}(\mathbf{X}_i) G_{\bar{C}(K)}(\mathbf{X}_i)}$$

We say that $\hat{\Psi}^{\mathsf{IPTW}}$ is (asymptotically) linear with influence function

$$\phi^{\mathsf{IPTW}}(P)(\mathbf{X}) = Y(K) \frac{\mathbb{1}(\bar{A}(K-1) = \bar{a}(K-1), \bar{C}(K) = 1)}{G_{\bar{A}(K-1)}(\mathbf{X}) G_{\bar{C}(K)}(\mathbf{X})} - \Psi(P)$$

Influence functions

If $G_{\bar{A}(k)}(\mathbf{X}) = \prod_{i=0}^{k} G_{A(i)}(\mathbf{X})$, $G_{\bar{C}(k)}(\mathbf{X}) = \prod_{i=1}^{k} G_{C(i)}(\mathbf{X})$, we know that

$$\Psi_0 = \Psi(P_{\mathbf{X}}) = \mathbb{E}_{P_{\mathbf{X}}}\left\{ Y(K) \frac{\mathbb{1}(\bar{A}(K-1) = \bar{a}(K-1), \bar{C}(K) = 1)}{G_{\bar{A}(K-1)}(\mathbf{X}) G_{\bar{C}(K)}(\mathbf{X})} \right\}$$

Replacing $\mathbb{E}_{P_{\mathbf{X}}}$ with the empirical average yields the estimator

$$\hat{\Psi}^{\mathsf{IPTW}} = \frac{1}{n} \sum_{i=1}^{n} Y_i(K) \frac{\mathbb{1}(\bar{A}_i(K-1) = \bar{a}(K-1), \bar{C}_i(K) = 1)}{G_{\bar{A}(K-1)}(\mathbf{X}_i) G_{\bar{C}(K)}(\mathbf{X}_i)}$$

We say that $\hat{\Psi}^{\mathsf{IPTW}}$ is (asymptotically) linear with influence function

$$\phi^{\mathsf{IPTW}}(P)(\mathbf{X}) = Y(K) \frac{\mathbb{1}(\bar{A}(K-1) = \bar{a}(K-1), \bar{C}(K) = 1)}{G_{\bar{A}(K-1)}(\mathbf{X}) G_{\bar{C}(K)}(\mathbf{X})} - \Psi(P)$$

and see that $\sqrt{n}(\hat{\Psi}^{\mathsf{IPTW}} - \Psi_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathbb{E}\{\phi^{\mathsf{IPTW}}(P_{\mathbf{X}})(\mathbf{X})^2\})$.

The efficient influence function

There are (potentially) many different estimators $\hat{\Psi}$ of Ψ_0 such that

$$\hat{\Psi} - \Psi_0 = \frac{1}{n} \sum_{i=1}^{n} \phi(P_{\mathbf{X}})(\mathbf{X}_i) + o_P(n^{-\frac{1}{2}})$$

for some influence function ϕ. Each is asymptotically normal with variance $\mathbb{E}\{\phi(P_{\mathbf{X}})(\mathbf{X})^2\}$.

The efficient influence function

There are (potentially) many different estimators $\hat{\Psi}$ of Ψ_0 such that

$$\hat{\Psi} - \Psi_0 = \frac{1}{n}\sum_{i=1}^{n} \phi(P_{\mathbf{X}})(\mathbf{X}_i) + o_P(n^{-\frac{1}{2}})$$

for some influence function ϕ. Each is asymptotically normal with variance $\mathbb{E}\{\phi(P_{\mathbf{X}})(\mathbf{X})^2\}$.

$\rightarrow$ There is some ϕ^* such that $\mathbb{E}\{\phi(P_{\mathbf{X}})(\mathbf{X})^2\} \geq \mathbb{E}\{\phi^*(P_{\mathbf{X}})(\mathbf{X})^2\}$ for all other ϕ. We call ϕ^* the efficient influence function (EIF).

The efficient influence function

There are (potentially) many different estimators $\hat{\Psi}$ of Ψ_0 such that

$$\hat{\Psi} - \Psi_0 = \frac{1}{n}\sum_{i=1}^{n} \phi(P_{\mathbf{X}})(\mathbf{X}_i) + o_P(n^{-\frac{1}{2}})$$

for some influence function ϕ. Each is asymptotically normal with variance $\mathbb{E}\{\phi(P_{\mathbf{X}})(\mathbf{X})^2\}$.

$\rightarrow$ There is some ϕ^* such that $\mathbb{E}\{\phi(P_{\mathbf{X}})(\mathbf{X})^2\} \geq \mathbb{E}\{\phi^*(P_{\mathbf{X}})(\mathbf{X})^2\}$ for all other ϕ. We call ϕ^* the efficient influence function (EIF).

For details on influence functions see e.g. Kennedy (2016).

The efficient influence function for our problem

It is possible to derive the EIF for our problem (van der Laan and Gruber, 2012). It consists of one component for each time point k:

$$\phi^*(P)(\mathbf{X}) = \sum_{k=0}^{K} \phi_k^*(P)(\mathbf{X})$$

$$\phi_K^*(P)(\mathbf{X}) = \frac{\mathbb{1}(\bar{A}(K-1) = \bar{a}(K-1), \bar{C}(K) = 1)}{G_{\bar{A}(K-1)}(\mathbf{X}) G_{\bar{C}(K)}(\mathbf{X})} (Y(K) - \bar{Q}^{\bar{a}}_{L(K)})$$

$$\phi_k^*(P)(\mathbf{X}) = \frac{\mathbb{1}(\bar{A}(k-1) = \bar{a}(k-1), \bar{C}(k) = 1))}{G_{\bar{A}(k-1)}(\mathbf{X}) G_{\bar{C}(k)}(\mathbf{X})} (\bar{Q}^{\bar{a}}_{L(k+1)} - \bar{Q}^{\bar{a}}_{L(k)})$$

$$\phi_0^*(P)(\mathbf{X}) = \bar{Q}^{\bar{a}}_{L(1)} - \bar{Q}^{\bar{a}}_{L(0)} = \bar{Q}^{\bar{a}}_{L(1)} - \Psi(P)$$

Two approaches to efficient estimation

We now have the tools to construct efficient estimators. Two approaches (not exhaustive):

Two approaches to efficient estimation

We now have the tools to construct efficient estimators. Two approaches (not exhaustive):

Change estimator: Instead of using $\hat{\Psi}^{\mathsf{SR}}$ or $\hat{\Psi}^{\mathsf{IPTW}}$, we could construct a new estimator based on the EIF:

$$\hat{\Psi}^{\mathsf{EIF}} = \frac{1}{n}\sum_{i=1}^{n}(\phi^*(\hat{P}_{\mathbf{X}})(\mathbf{X_i}) + \Psi_0)$$

where we take the empirical mean of the stochastic part of the EIF, plugging in an estimate $\hat{P}_{\mathbf{X}}$ of $P_{\mathbf{X}}$ (or more precisely, of G_A, G_C, $\bar{Q}^{\bar{a}}$).

Two approaches to efficient estimation

We now have the tools to construct efficient estimators. Two approaches (not exhaustive):

Change estimator: Instead of using $\hat{\Psi}^{\mathsf{SR}}$ or $\hat{\Psi}^{\mathsf{IPTW}}$, we could construct a new estimator based on the EIF:

$$\hat{\Psi}^{\mathsf{EIF}} = \frac{1}{n}\sum_{i=1}^{n}(\phi^*(\hat{P}_{\mathbf{X}})(\mathbf{X_i}) + \Psi_0)$$

where we take the empirical mean of the stochastic part of the EIF, plugging in an estimate $\hat{P}_{\mathbf{X}}$ of $P_{\mathbf{X}}$ (or more precisely, of $G_A, G_C, \bar{Q}^{\bar{a}}$).

Change estimation of nuisance parameters: Instead of changing the estimator $\hat{\Psi}^{\mathsf{SR}} = \Psi^{\mathsf{SR}}(\hat{\bar{Q}}^{\bar{a}})$, we change the estimator $\hat{\bar{Q}}^{\bar{a}}$ into some $\hat{\bar{Q}}^{\bar{a},*}$ such that $\Psi^{\mathsf{SR}}(\hat{\bar{Q}}^{\bar{a},*})$ becomes efficient $\rightarrow$ the TMLE approach.

The targeting step

The TMLE approach updates the initial estimate $\hat{\bar{Q}}^{\bar{a}}$ into $\hat{\bar{Q}}^{\bar{a},*}$ with the help of the EIF. This is the targeting step. Involves two ingredients:

The targeting step

The TMLE approach updates the initial estimate $\hat{\bar{Q}}^{\bar{a}}$ into $\hat{\bar{Q}}^{\bar{a},*}$ with the help of the EIF. This is the targeting step. Involves two ingredients:

1. A loss function $\mathcal{L}(\bar{Q}^{\bar{a}})(\mathbf{X})$.

The targeting step

The TMLE approach updates the initial estimate $\hat{\bar{Q}}^{\bar{a}}$ into $\hat{\bar{Q}}^{\bar{a},*}$ with the help of the EIF. This is the targeting step. Involves two ingredients:

1. A loss function $\mathcal{L}(\bar{Q}^{\bar{a}})(\mathbf{X})$.
2. A parametric fluctuation model $\{\bar{Q}^{\bar{a}}(\epsilon)\}$ indexed by $\epsilon \in \mathbb{R}$.

The targeting step

The TMLE approach updates the initial estimate $\hat{\bar{Q}}^{\bar{a}}$ into $\hat{\bar{Q}}^{\bar{a},*}$ with the help of the EIF. This is the targeting step. Involves two ingredients:

1. A loss function $\mathcal{L}(\bar{Q}^{\bar{a}})(\mathbf{X})$.
2. A parametric fluctuation model $\{\bar{Q}^{\bar{a}}(\epsilon)\}$ indexed by $\epsilon \in \mathbb{R}$.

The ingredients must combine in a certain way:

$$\frac{d}{d\epsilon}\bigg|_{\epsilon=0} \mathcal{L}(\bar{Q}^{\bar{a}}(\epsilon))(\mathbf{X}) = \phi^*(P_{\mathbf{X}})(\mathbf{X})$$

The targeting step

The TMLE approach updates the initial estimate $\hat{\bar{Q}}^{\bar{a}}$ into $\hat{\bar{Q}}^{\bar{a},*}$ with the help of the EIF. This is the targeting step. Involves two ingredients:

1. A loss function $\mathcal{L}(\bar{Q}^{\bar{a}})(\mathbf{X})$.
2. A parametric fluctuation model $\{\bar{Q}^{\bar{a}}(\epsilon)\}$ indexed by $\epsilon \in \mathbb{R}$.

The ingredients must combine in a certain way:

$$\frac{d}{d\epsilon}\bigg|_{\epsilon=0} \mathcal{L}(\bar{Q}^{\bar{a}}(\epsilon))(\mathbf{X}) = \phi^*(P_{\mathbf{X}})(\mathbf{X})$$

We then make the update $\hat{\bar{Q}}^{\bar{a},*} := \hat{\bar{Q}}^{\bar{a}}(\hat{\epsilon})$

$$\hat{\epsilon} := \operatorname*{argmin}_{\epsilon} \frac{1}{n} \sum_{i=1}^{n} \mathcal{L}(\hat{\bar{Q}}^{\bar{a}}(\epsilon))(\mathbf{X}_i)$$

The targeting step for our problem

Due to the decomposition $\phi^*(P)(\mathbf{X}) = \sum_{k=0}^{K} \phi_k^*(P)(\mathbf{X})$, we can formulate the targeting step one k at a time and use a sequential targeting procedure.

The targeting step for our problem

Due to the decomposition $\phi^*(P)(\mathbf{X}) = \sum_{k=0}^{K} \phi_k^*(P)(\mathbf{X})$, we can formulate the targeting step one k at a time and use a sequential targeting procedure. We choose loss function

$$\mathcal{L}_{k,\bar{Q}^{\bar{a}}_{L(k+1)}}(\bar{Q}^{\bar{a}}_{L(k)}) = -(\bar{Q}^{\bar{a}}_{L(k+1)} \log(\bar{Q}^{\bar{a}}_{L(k)}) + (1-\bar{Q}^{\bar{a}}_{L(k+1)}) \log(1-\bar{Q}^{\bar{a}}_{L(k)}))$$

The targeting step for our problem

Due to the decomposition $\phi^*(P)(\mathbf{X}) = \sum_{k=0}^{K} \phi_k^*(P)(\mathbf{X})$, we can formulate the targeting step one k at a time and use a sequential targeting procedure. We choose loss function

$$\mathcal{L}_{k,\bar{Q}^{\bar{a}}_{L(k+1)}}(\bar{Q}^{\bar{a}}_{L(k)}) = -(\bar{Q}^{\bar{a}}_{L(k+1)} \log(\bar{Q}^{\bar{a}}_{L(k)}) + (1-\bar{Q}^{\bar{a}}_{L(k+1)}) \log(1-\bar{Q}^{\bar{a}}_{L(k)}))$$

and fluctuation model

$$\operatorname{logit}(\bar{Q}^{\bar{a}}_{L(k)}(\epsilon_k)) = \operatorname{logit}(\bar{Q}^{\bar{a}}_{L(k)}) + \epsilon_k \frac{\mathbb{1}(\bar{A}(k-1) = \bar{a}(k-1), \bar{C}(k) = 1)}{G_{\bar{A}(k-1)}(\mathbf{X})\, G_{\bar{C}(k)}(\mathbf{X})}$$

The targeting step for our problem

Due to the decomposition $\phi^*(P)(\mathbf{X}) = \sum_{k=0}^{K} \phi_k^*(P)(\mathbf{X})$, we can formulate the targeting step one k at a time and use a sequential targeting procedure. We choose loss function

$$\mathcal{L}_{k,\bar{Q}^{\bar{a}}_{L(k+1)}}(\bar{Q}^{\bar{a}}_{L(k)}) = -(\bar{Q}^{\bar{a}}_{L(k+1)} \log(\bar{Q}^{\bar{a}}_{L(k)}) + (1 - \bar{Q}^{\bar{a}}_{L(k+1)}) \log(1 - \bar{Q}^{\bar{a}}_{L(k)}))$$

and fluctuation model

$$\operatorname{logit}(\bar{Q}^{\bar{a}}_{L(k)}(\epsilon_k)) = \operatorname{logit}(\bar{Q}^{\bar{a}}_{L(k)}) + \epsilon_k \frac{\mathbb{1}(\bar{A}(k-1) = \bar{a}(k-1), \bar{C}(k) = 1)}{G_{\bar{A}(k-1)}(\mathbf{X}) G_{\bar{C}(k)}(\mathbf{X})}$$

Plug $\bar{Q}^{\bar{a}}_{L(k)}(\epsilon_k)$ into loss and minimize sample loss over ϵ_k numerically.

Update as logistic regression with clever covariate

Minimizing the loss over ϵ_k is equivalent to doing logistic regression of $\bar{Q}^{\bar{a}}_{L(k+1)}$ with the IPTW:

$$\frac{\mathbb{1}(\bar{A}(k-1) = \bar{a}(k-1), \bar{C}(k) = 1)}{G_{\bar{A}(k-1)}(\mathbf{X}) G_{\bar{C}(k)}(\mathbf{X})}$$

as a so called clever covariate, and $\text{logit}(\bar{Q}^{\bar{a}}_{L(k)})$ as offset.

NB: No intercept term!

This is just a clever way of minimizing the loss over ϵ_k.

Sequential regression with targeting step I

First estimate all treatment/censoring propensity scores and get predictions $\hat{G}_{A(k)}$ and $\hat{G}_{C(k)}$.

Sequential regression with targeting step I

First estimate all treatment/censoring propensity scores and get predictions $\hat{G}_{A(k)}$ and $\hat{G}_{C(k)}$.

1. Fit model with $Y(K)$ as outcome, conditional on the past of $Y(K)$ $\rightarrow$ Predictions $\hat{\hat{Q}}^{\bar{a}}_{L(K)}$.

Sequential regression with targeting step I

First estimate all treatment/censoring propensity scores and get predictions $\hat{G}_{A(k)}$ and $\hat{G}_{C(k)}$.

1. Fit model with $Y(K)$ as outcome, conditional on the past of $Y(K)$ $\rightarrow$ Predictions $\hat{\hat{Q}}^{\bar{a}}_{L(K)}$.
   1.1 Fit logistic model using $Y(K)$ as outcome, covariate $\frac{\mathbb{1}(\bar{A}(K-1)=\bar{a}(K-1),\bar{C}(K)=1)}{\hat{G}_{\bar{A}(K-1)}(\mathbf{X})\hat{G}_{\bar{C}(K)}(\mathbf{X})}$ and offset $\operatorname{logit}(\hat{\hat{Q}}^{\bar{a}}_{L(K)})$, no intercept.
   $\rightarrow$ Targeted predictions $\hat{\hat{Q}}^{\bar{a},*}_{L(K)}$.

Sequential regression with targeting step I

First estimate all treatment/censoring propensity scores and get predictions $\hat{G}_{A(k)}$ and $\hat{G}_{C(k)}$.

1. Fit model with $Y(K)$ as outcome, conditional on the past of $Y(K)$ $\rightarrow$ Predictions $\hat{\hat{Q}}^{\bar{a}}_{L(K)}$.
   1.1 Fit logistic model using $Y(K)$ as outcome, covariate $\frac{\mathbb{1}(\bar{A}(K-1)=\bar{a}(K-1),\bar{C}(K)=1)}{\hat{G}_{\bar{A}(K-1)}(\mathbf{X})\hat{G}_{\bar{C}(K)}(\mathbf{X})}$ and offset $\operatorname{logit}(\hat{\hat{Q}}^{\bar{a}}_{L(K)})$, no intercept.
   $\rightarrow$ Targeted predictions $\hat{\hat{Q}}^{\bar{a},*}_{L(K)}$.
2. Fit model using $\hat{\hat{Q}}^{\bar{a},*}_{L(K)}$ as outcome, conditional on the past of $Y(K-1)$ $\rightarrow$ Predictions $\hat{\hat{Q}}^{\bar{a}}_{L(K-1)}$.

Sequential regression with targeting step I

First estimate all treatment/censoring propensity scores and get predictions $\hat{G}_{A(k)}$ and $\hat{G}_{C(k)}$.

1. Fit model with $Y(K)$ as outcome, conditional on the past of $Y(K)$ $\rightarrow$ Predictions $\hat{\hat{Q}}^{\bar{a}}_{L(K)}$.
   1.1 Fit logistic model using $Y(K)$ as outcome, covariate $\frac{\mathbb{1}(\bar{A}(K-1)=\bar{a}(K-1),\bar{C}(K)=1)}{\hat{G}_{\bar{A}(K-1)}(\mathbf{X})\hat{G}_{\bar{C}(K)}(\mathbf{X})}$ and offset $\mathrm{logit}(\hat{\hat{Q}}^{\bar{a}}_{L(K)})$, no intercept.
   $\rightarrow$ Targeted predictions $\hat{\hat{Q}}^{\bar{a},*}_{L(K)}$.
2. Fit model using $\hat{\hat{Q}}^{\bar{a},*}_{L(K)}$ as outcome, conditional on the past of $Y(K-1)$ $\rightarrow$ Predictions $\hat{\hat{Q}}^{\bar{a}}_{L(K-1)}$.
   2.1 Fit logistic model using $\hat{\hat{Q}}^{\bar{a},*}_{L(K)}$ as outcome, covariate $\frac{\mathbb{1}(\bar{A}(K-2)=\bar{a}(K-2),\bar{C}(K-1)=1)}{\hat{G}_{\bar{A}(K-2)}(\mathbf{X})\hat{G}_{\bar{C}(K-1)}(\mathbf{X})}$ and offset $\mathrm{logit}(\hat{\hat{Q}}^{\bar{a}}_{L(K-1)})$, no intercept.
   $\rightarrow$ Targeted predictions $\hat{\hat{Q}}^{\bar{a},*}_{L(K-1)}$.

Sequential regression with targeting step II

3. Repeat until targeted predictions $\hat{\hat{Q}}^{\bar{a},*}_{L(1)}$.

Sequential regression with targeting step II

3. Repeat until targeted predictions $\hat{\hat{Q}}^{\bar{a},*}_{L(1)}$.
4. Compute $\hat{\Psi}^{\mathsf{LTMLE}} = \hat{\hat{Q}}^{\bar{a},*}_{L(0)} = \frac{1}{n}\sum_{i=1}^{n} \hat{\hat{Q}}^{\bar{a},*}_{L(1),i}$.

Sequential regression with targeting step II

3. Repeat until targeted predictions $\hat{\bar{Q}}^{\bar{a},*}_{L(1)}$.
4. Compute $\hat{\Psi}^{\mathsf{LTMLE}} = \hat{\bar{Q}}^{\bar{a},*}_{L(0)} = \frac{1}{n}\sum_{i=1}^{n} \hat{\bar{Q}}^{\bar{a},*}_{L(1),i}$.

The result is the (L)TMLE of our target parameter.

Targeting step in R: Setup

Same data structure as earlier:

head(d)

	L_0	A_0	C_1	Y_1	L_1	A_1	C_2	Y_2
1 :	1	1	1	0	1	0	1	0
2 :	0	0	1	1	1	0	1	1
3 :	1	0	1	0	1	1	0	0

```
d1 <- copy(d)[, A_0 := 1][, A_1 := 1]
```

Targeting step in R: Treatment and censoring propensity models

First estimate all treatment and censoring propensities:

```
G_A0 <- glm(A_0 ~ L_0, family = binomial(), data = d)
G_A0_hat <- predict(G_A0, newdata = d1, type = "response")

G_A1 <- glm(A_1 ~ L_0 + A_0 + L_1, family = binomial(),
            data = d[C_1 == 1 & Y_1 == 0])
G_A1_hat <- predict(G_A1, newdata = d1, type = "response")

G_C1 <- glm(C_1 ~ L_0 + A_0, family = binomial(), data = d)
G_C1_hat <- predict(G_C1, newdata = d1, type = "response")

G_C2 <- glm(C_2 ~ L_0 + A_0 + L_1 + A_1, family = binomial(),
            data = d[C_1 == 1 & Y_1 == 0])
G_C2_hat <- predict(G_C2, newdata = d1, type = "response")

d[, G_A0_hat := G_A0_hat][, G_A1_hat := G_A1_hat][
    , G_C1_hat := G_C1_hat][, G_C2_hat := G_C2_hat]
```

Targeting step in R: Time 2

Fit outcome regression as before:

```
Q2 <- glm(Y_2 ~ L_0 + A_0 + L_1 + A_1,
          family = binomial(),
          data = d[Y_1 == 0 & C_1 == 1 & C_2 == 1])
Q2_hat <- predict(Q2, newdata = d1, type = "response")
d[, Q2_hat := Y_1 + (1 - Y_1) * Q2_hat]
```

Targeting step in R: Time 2

Fit outcome regression as before:

```
Q2 <- glm(Y_2 ~ L_0 + A_0 + L_1 + A_1,
          family = binomial(),
          data = d[Y_1 == 0 & C_1 == 1 & C_2 == 1])
Q2_hat <- predict(Q2, newdata = d1, type = "response")
d[, Q2_hat := Y_1 + (1 - Y_1) * Q2_hat]
```

Do the targeting step:

```
Q2star <- glm(Y_2 ~ I((A_0 == 1 & A_1 == 1) /
              (G_A0_hat * G_A1_hat * G_C1_hat * G_C2_hat)) - 1,
              family = binomial(), offset = qlogis(Q2_hat),
              data = d[Y_1 == 0 & C_1 == 1 & C_2 == 1])
Q2star_hat <- predict(Q2star, newdata = d1,
                      type = "response")
d[, Q2star_hat := Y_1 + (1 - Y_1) * Q2star_hat]
```

Targeting step in R: Time 1

We use targeted estimate from previous step as outcome:

```
Q1 <- glm(Q2star_hat ~ L_0 + A_0, family = quasibinomial(),
          data = d[C_1 == 1])
Q1_hat <- predict(Q1, newdata = d1, type = "response")
d[, Q1_hat := Q1_hat]
```

Targeting step in R: Time 1

We use targeted estimate from previous step as outcome:

```
Q1 <- glm(Q2star_hat ~ L_0 + A_0, family = quasibinomial(),
          data = d[C_1 == 1])
Q1_hat <- predict(Q1, newdata = d1, type = "response")
d[, Q1_hat := Q1_hat]
```

And then do another targeting step:

```
Q1star <- glm(Q2star_hat ~
                 I((A_0 == 1) / (G_A0_hat * G_C1_hat)) - 1,
              family = quasibinomial(),
              offset = qlogis(Q1_hat),
              data = d[C_1 == 1])
Q1star_hat <- predict(Q1star, newdata = d1,
                      type = "response")
```

Targeting step in R: Results

Again take sample mean for the final estimate:

```
mean(Q1star_hat)
[1] 0.04849994

fit <- ltmle::ltmle(
  d, # have to use original data without intermediate calcs
  Anodes = c("A_0", "A_1"),
  Cnodes = c("C_1", "C_2"),
  Lnodes = c("L_0", "L_1"),
  Ynodes = c("Y_1", "Y_2"),
  survivalOutcome = TRUE,
  abar = c(1, 1),
  gcomp = FALSE
)
[1] 0.04849969
```

Discrepancy (hopefully) due to implementation details, e.g. bounding of weights in ltmle.

References

- Kennedy, Edward H. (2016). *Semiparametric theory and empirical processes in causal inference*. arXiv: 1510.04740 [math.ST].
- Petersen, Maya, Joshua Schwab, Susan Gruber, Nello Blaser, Michael Schomaker, and Mark van der Laan (2014). "Targeted Maximum Likelihood Estimation for Dynamic and Static Longitudinal Marginal Structural Working Models". In: *Journal of Causal Inference* 2.2, pp. 147–185. ISSN: 2193-3677. DOI: 10.1515/jci-2013-0007.
- van der Laan, Mark J. and Susan Gruber (2012). "Targeted Minimum Loss Based Estimation of Causal Effects of Multiple Time Point Interventions". In: *The International Journal of Biostatistics* 8.1. DOI: doi:10.1515/1557-4679.1370.