Application of Time-to-Event Methods in the Assessment of Safety in Clinical Trials

Progress, Updates, Problems

William Jen Hoe Koh

May 9, 2013

Overview

- Marginal vs Conditional
- What is TMLE?
- Key Estimation procedure
- Some test simulations
- Hidden assumptions
- Hidden Statistical Properties

Brief Review: Time-to-event outcomes

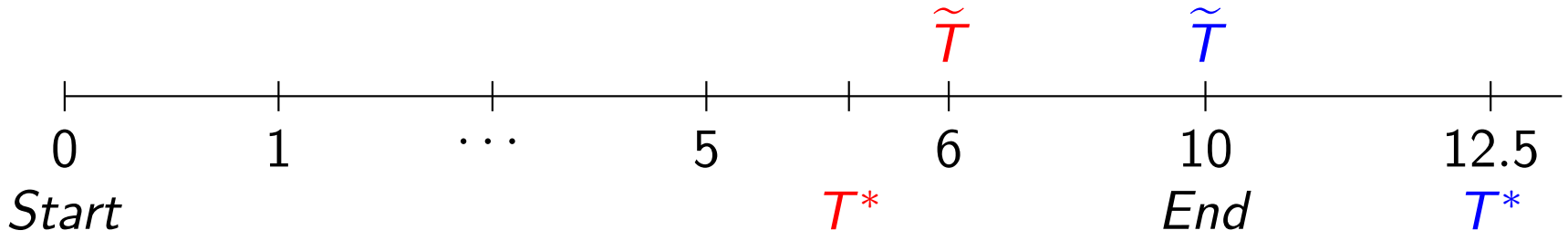
$$S(t) = Pr(T > t) = 1 - F(t)$$

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{\Pr[t \leq T < t + \Delta t | T \geq t]}{\Delta t}$$

Event of Interest: Infection/AE at clinic visit

Right censored data: Non-ignorable missing data.

Discrete Failure Time (Zhang and Gilbert, 2010)



$$\lambda(t_j) = \Pr(T = t_j | T > t_{j-1})$$

T^*: "True" failure time (Unobserved due to discrete follow-up)

$T = t_j$ if $T^* \in [t_{j-1}, t_j)$ with t_j for $j = 1, \cdots, 10$.

$\widetilde{T} = min(T, C)$: where C is our censoring time.

$\Delta = I(T \leq C)$: Indicator of subject **not** being censored

Objectives (Moore and van der Laan, 2009b)

- Estimate marginal treatment specific survival at a fixed end point
- Use covariates to gain efficiency
- Provide a consistent estimator in the presence of informative censoring

Question we want to address

Approximate Scientific question: $\Pr(T_1 > t_k) - \Pr(T_0 > t_k)$

Event of Interest: First record of adverse event reported

Simplest analysis: Kaplan Meier Survival curves

Assumptions: Random censoring

Adjusted analysis: Cox-PH or Logistic regression (More assumptions)

RCT: Marginal or Conditional

Scientific question: $\Pr(T > t_k | A = 1) - \Pr(T > t_k | A = 0)$

Marginal

$P(T > t | A = a) = S_0(t | A = a)$

Estimated probability of survival past time t for treatment a for the entire population.

Conditional

$P(T > t | A = a, W) = S_0(t | A = a, W)$

Estimated probability of survival past time t for treatment a while holding W fixed.

What does TMLE estimate

Let $\Psi : \mathcal{M} \rightarrow \mathbb{R}$ is pathwise differentiable at any density $p_0 \in \mathcal{M}$

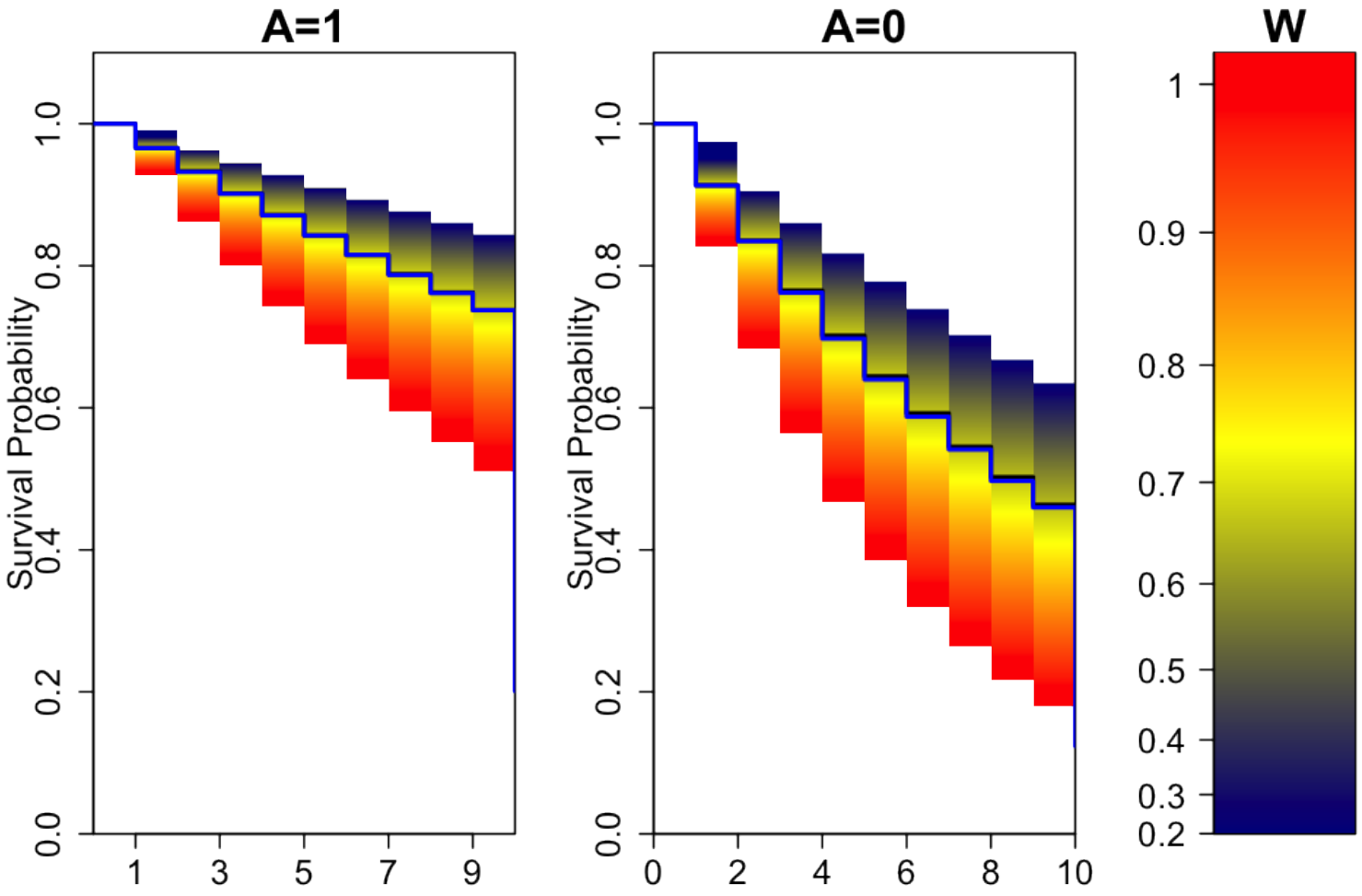
$$\Psi_1(p_0)(t_k) = \overbrace{\Pr(T_1 > t_k)}^{Marginal} = E_W[\overbrace{S_0(t_k|A=1,\mathbf{W})}^{Conditional}]$$
$$\Psi_0(p_0)(t_k) = \Pr(T_0 > t_k) = E_W[S_0(t_k|A=0,\mathbf{W}])$$
$$\Psi_{AD}(p_0)(t_k) = \Psi_1(p_0)(t_k) - \Psi_0(p_0)(t_k)$$

Consider the treatment group $A = 1$,

$Pr(T_1 > t_k|\mathbf{W}) = S_0(t_k|A=1,\mathbf{W})$: Probability of surviving beyond t_k when treatment is 1 given the covariates W.

$\Pr(T_1 > t_k) = E[S_0(t_k|A=1,\mathbf{W})]$: Probability of surviving beyond t_k when treatment is 1 averaging over the covariates W.

Example: Random treatment assignment of 0.5



$W \sim U(0.2, 1.2)$

$\lambda(t|A, W) = expit(-3 - A + W^2)I(t < 10) + I(t = 10)$

Idea behind TMLE (Van Der Laan, 2011)

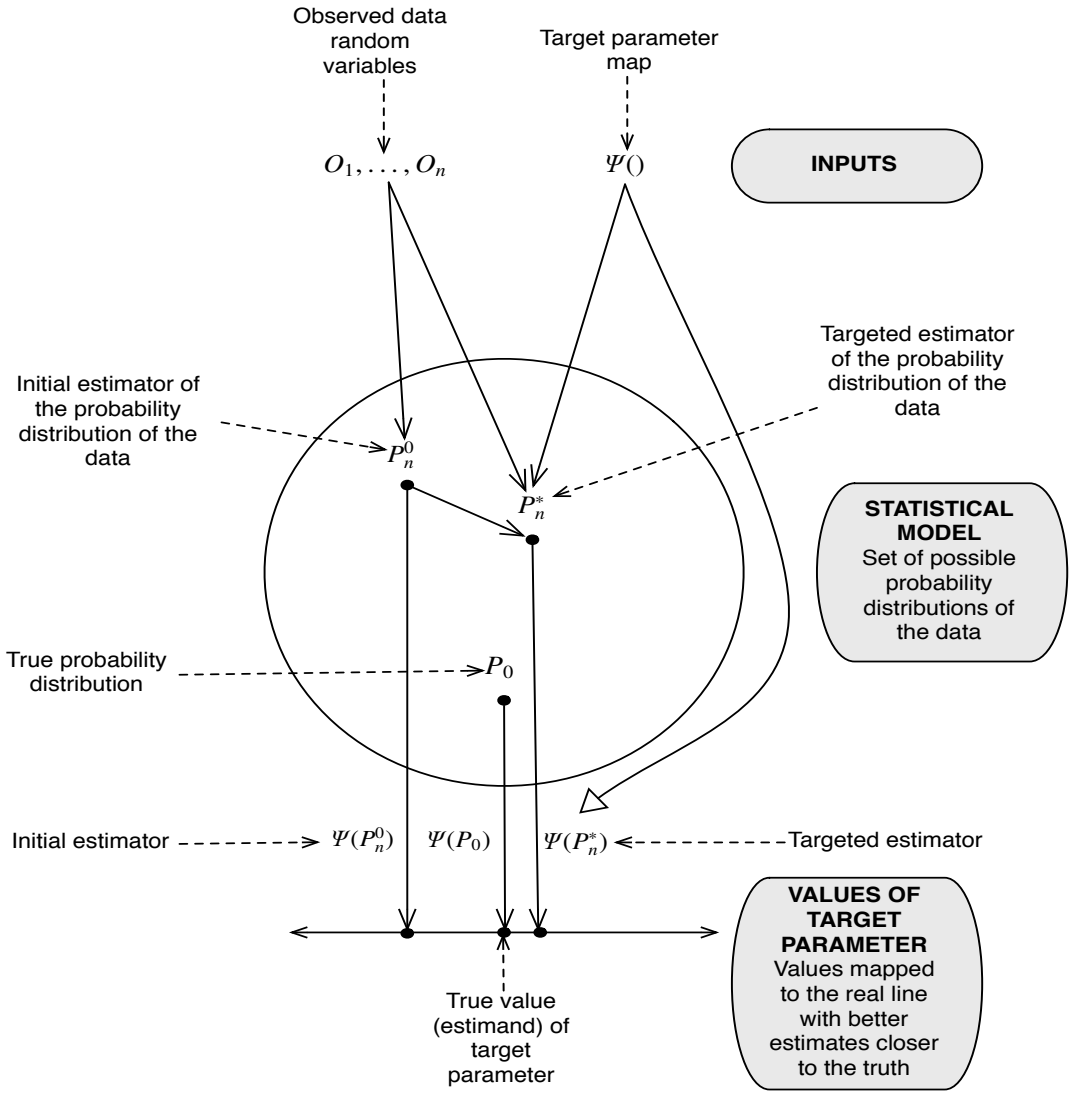


Fig. 5.1 TMLE flow chart.

Likelihood representation for 1 observation

$\mathbf{W}$: Covariates

A: Treatment

$(\widetilde{T}, \Delta)$: Time and event indicator

$\bar{G}(t_-|A, \mathbf{W}) = \Pr(C \geq t|A, \mathbf{W})$(Moore and van der Laan, 2009a)

$S(t_k|A, W) = \prod_{t \leq t_k}[1 - \lambda(t|A, W)]$

$$P_0(\mathcal{O}) = \Pr(\mathbf{W}, A, \widetilde{T}, \Delta) = \Pr(\widetilde{T}, \Delta|\mathbf{W}, A)\Pr(\mathbf{W})\Pr(A|\mathbf{W})$$

$$= \overbrace{\left[\lambda(t_k|A, \mathbf{W}) \prod_{t=1}^{t_k-1}(1 - \lambda(t_k|A, \mathbf{W}))\right]^{\delta} \left[\prod_{t=1}^{t_k}(1 - \lambda(t_k|A, \mathbf{W}))\right]^{1-\delta}}^{Q_{20}}$$

$$\underbrace{\bar{G}(t_-|A, \mathbf{W})^{\delta}\Pr(C = t_k|A, \mathbf{W})^{1-\delta}}_{g_{20}} \underbrace{\Pr(\mathbf{W})}_{Q_{10}} \underbrace{\Pr(A|\mathbf{W})}_{g_{10}}$$

Key Results for TMLE: Doubly robust

Q_0

Q_{10}: Distribution of the baseline covariates

Q_{20}: Conditional distribution of the hazard given treatment and baseline covariates

g_0

g_{10}: Treatment mechanism

g_{20}: Conditional distribution of the censoring distribution given treatment and baseline covariates

Either Q_0 or g_0 is correct, then TMLE is consistent

TMLE as Plug-in estimators

1. Estimate the $\hat{g}^0(A=1|W) = \frac{1}{n}\sum_{i=1}^{n} A_i$
2. Estimate the conditional probability of censoring $\bar{G}^0(t_-|A,W)$
3. Estimate the $\hat{\lambda}^0(t|A,W)$ via logistic regression

$$\text{logit}[\hat{\lambda}^0(t|A,W)] = \sum_{i=1}^{K} \alpha_i I(t=i) + \beta_A A + \boldsymbol{\beta}_W W$$

4. Update the above model using $\hat{\lambda}^0(t|A,W)$ by including the penalty $\epsilon = \{\epsilon_0, \epsilon_1\}$ and the "clever" covariate $\hat{h}^0(t,A,W) = \{\hat{h}_0, \hat{h}_1\}$

$$\text{logit}[\hat{\lambda}^1(t|A,W)] = \text{logit}[\hat{\lambda}^0(t|A,W)] + \epsilon^T \hat{h}^0(t,A,W)$$

$$\widehat{h}_i(t,A,W) = -\frac{I(A=i)I(t \leq t_k)}{\hat{g}^0(A=i|W)\bar{G}^0(t_-|A,W)} \frac{\widehat{S}(t_k|A,W)}{\widehat{S}(t|A,W)}$$

TMLE as Plug-in estimators continued

5. Use current estimate of $\widehat{\lambda}^1(t|A, W)$ to update $\hat{h}^1(t, A, W)$

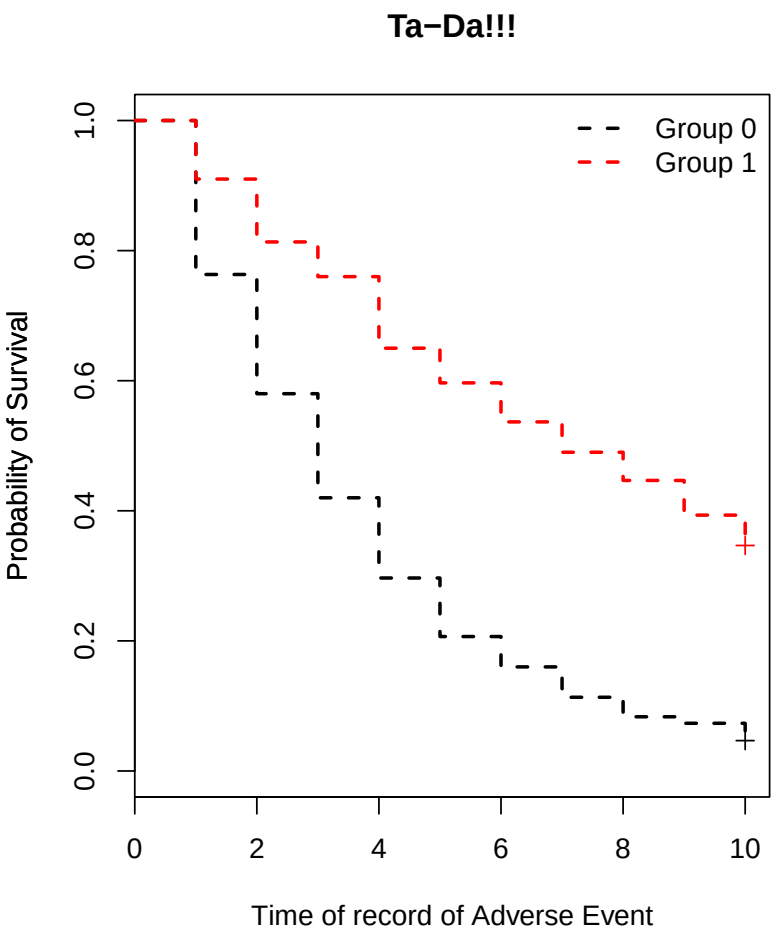
$$\widehat{S}(t_k|A, W) = \prod_{t \leq t_k} [1 - \widehat{\lambda}^1(t|A, W)]$$

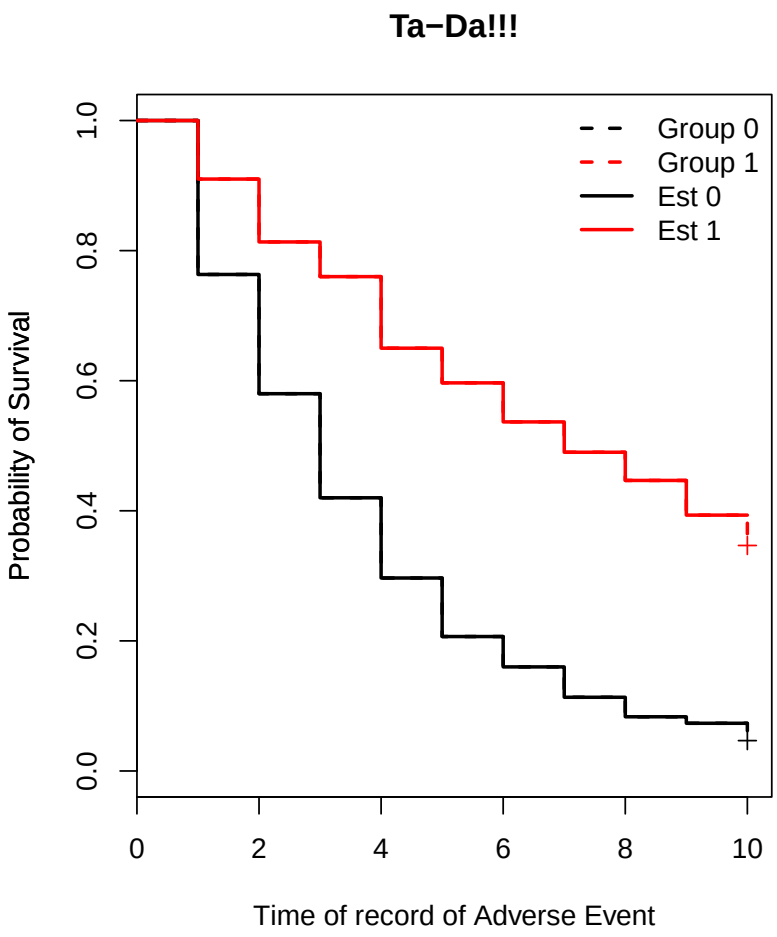
6. Iterate 4 & 5 until $\widehat{\epsilon} \rightarrow 0$
7. Plug-in final estimate of $\widehat{S}_i^*(t_k|A = i, W)$ for $i = 0, 1$

$$\widehat{\Psi}_1(p_0)(t_k) = \frac{1}{n} \sum_{i=1}^{n} \widehat{S}^*(t_k|A = 1, W_i)$$

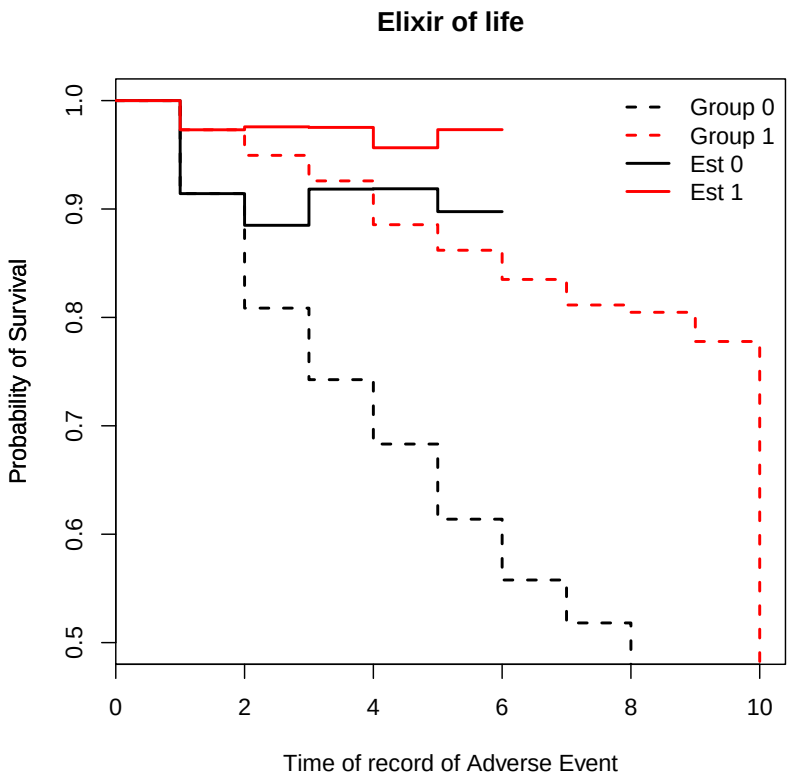
$$\widehat{\Psi}_0(p_0)(t_k) = \frac{1}{n} \sum_{i=1}^{n} \widehat{S}^*(t_k|A = 0, W_i)$$

Test simulations with no covariates





Test Simulations for weak covariates



Simulations

$W \sim U(0.2, 1.2)$; $A \sim Bin(0, \frac{1}{2})$

$\lambda(t|A, W) = expit(-3 - A + W^2)I(t < 10) + I(t = 10)$

Problem

Survival function never increases over time.

Acknowledgements: Taken from bartsblackboard.com

Next steps

Compute the variance/Bootstrap

Rerun simulations

Fill in the gaps in the paper

Extensions

References

Moore, K. and van der Laan, M. (2009a). Increasing power in randomized trials with right censored outcomes through covariate adjustment. *Journal of biopharmaceutical statistics*, 19(6):1099–1131.

Moore, K. L. and van der Laan, M. J. (2009b). Application of time-to-event methods in the assessment of safety in clinical trials (chapter 20). In Peace, K. E., editor, *Design and Analysis of Clinical Trials with Time-to-Event Endpoints*, pages 455–482. Chapman and Hall/CRC; Biostatistics Series.

Stitelman, O. and van der Laan, M. (2010). Collaborative targeted maximum likelihood for time to event data. *The international journal of biostatistics*, 6(1):Article–21.

Van Der Laan, M. (2011). *Targeted Learning: Causal Inference for Observational and Experimental Data*. Springer.

Zhang, M. and Gilbert, P. B. (2010). Increasing the efficiency of prevention trials by incorporating baseline covariates. *Statistical communications in infectious diseases*, 2(1).

Thinking in counterfactuals

You take the red pill - you stay in Wonderland and I show you how deep the rabbit hole goes.

You take the blue pill - the story ends, you wake up in your bed and believe whatever you want to believe.

Morpheus, The Matrix

Hidden Assumptions: Thinking in counterfactuals

For an individual in the placebo arm...

$$IC^*_{0t_k}(p_O) = \sum_{t \leq t_k} [I(\tilde{T} = t, \Delta = 1) - I(\tilde{T} \geq t)\lambda(t|A = 0, W)]h_0(t, A, W)$$
$$+ S_0(t_k|A = 0, W) - \Psi_0(p_O)(t_k)$$

$$h_i(t, A, W) = -\frac{I(A = i)}{g(1)\bar{G}(t_-|A, W)} \frac{S(t_k|A, W)}{S(t|A, W)} I(t \leq t_k)$$

Key assumptions 1: Coarsening at random

Definition of CAR (Stitelman and van der Laan, 2010)

Coarsened data structures are data structures where the full data is not observed.

"Coarsening mechanism" is only a function of the full data, i.e. the data in which you would have seen all counterfactuals, through the observed data

Implications

$dP_0(O) = Q_0(O)g_0(O|X)$

Q_0 is the density associated with full data

g_0 contains the censoring and treatment mechanism.

Statistical definitions

$\mathcal{M}$ is the set of possible probability distribution of $\mathcal{O}$ with probability distribution $P_0 \in \mathcal{M}$ where $\mathcal{M}$ is dominated by common measure μ. Hence, density $p = \frac{dP_0}{d\mu}$

$O_1, \cdots, O_n$ are n i.i.d realizations of $\mathcal{O}$.

$O_1, \cdots, O_n$ can be represented by the empirical probability distribution P_n placing mass $1/n$ on each of the n observations.

Statistical properties

TMLE uses solves the efficient Influence curve.

IDEA: If we can linearize our estimator

Linearity of estimator

An estimator ψ_n is an asymptotically linear estimator of a parameter ψ if

$$\psi_n - \psi = \frac{1}{n}\sum_{i=1}^{n} IC_P(O_i) + o_P\left(\frac{1}{\sqrt{n}}\right)$$

where the influence curve $IC_P(O)$ has expectation 0 and finite variance i.e. $IC_P(O) \in L_0^2(P)$ and moreover this should hold for all $P \in \mathcal{M}$.

Efficient Influence curve

Asymptotic Distribution

The asymptotic distribution (via linearity of the influence curve)

$$\sqrt{n}\left(\hat{\psi}^*(t_k) - \Psi(p_0)(t_k)\right) \to N(0, \sigma^2)$$

Estimate σ^2 by empirical variance

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} IC_P(O_i)^2$$

Implications

Compute Wald-like confidence intervals $\hat{\psi}^*(t_k) \pm 1.96\frac{\hat{\sigma}^2}{\sqrt{n}}$