Spring 2016

Introduction to Targeted Learning

Laura Balzer, *University of California, Berkeley*

Introduction to Targeted Learning

Laura Balzer, PhD

Department of Biostatistics
Harvard T.H. Chan School of Public Health

December 2016

Outline

1. Motivating example
2. Roadmap for Targeted Learning
   - Scientific question $\rightarrow$ Causal parameter $\rightarrow$ Estimation procedure $\rightarrow$ Interpretation
3. Summary & Discussion

The SEARCH Trial

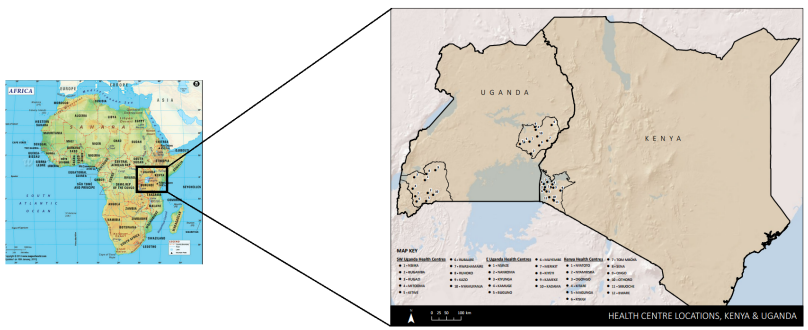
SEARCH

SUSTAINABLE EAST AFRICA RESEARCH IN COMMUNITY HEALTH

- Multinational, multidisciplinary consortium
- Led by Drs. Diane Havlir (UCSF), Moses Kamya (Makerere University) & Maya Petersen (UCB)
- Mission: End AIDS in East Africa
- `www.searchendaids.com`

The SEARCH Trial

- Six-year cluster randomized trial
- 32 communities in rural Uganda and Kenya
- $\approx$ 320,000 people
- Phase1: Early HIV diagnosis with immediate and streamlined ART (antiretroviral therapy)
- Phase2: Targeted PrEP (Pre-Exposure Prophylaxis), targeted HIV testing, and targeted care on top of universal and streamlined ART



The SEARCH Trial

Focus on Phase1

- Intervention: all HIV+ offered immediate ART with streamlined care
  - Services for initiation, linkage and retention
  - Annual, community-wide testing for HIV
- Control: all HIV+ offered ART according to in-country guidelines
- Primary outcome: three-year cumulative incidence of HIV



The SEARCH Trial

- At baseline in SEARCH, we sought to test all stable, adult residents for HIV
- Hybrid testing scheme:
  - Mobile community health campaigns (CHCs) offered HIV testing along with multi-disease prevention and treatment services
  - Home-based testing for those not attending a CHC
- Tested 131,307 of 146,906 adults in rural Uganda and Kenya
  - Achieved 89% testing coverage

We often ask causal questions

Some scientific possible questions:

- Who did we miss with the hybrid scheme?
  - Descriptive
- What are the risk factors "significantly" associated with not testing?
  - Descriptive
- What is the effect of increased mobility on the risk of not testing?
  - Causal

Causal Roadmap as a Tool

1. Scientific question
2. Causal model
3. Counterfactuals & causal parameter
4. Observed data & statistical model
5. Identifiability & statistical parameter
6. Estimation
7. Interpretation

1. Specify the scientific question

- What is the effect of increased mobility on the risk of not testing?
- How would the risk of not testing differ if all adults lived 1+ month away vs. <1 month away?
  - Inference about testing uptake under different conditions
- Many other possible causal questions possible

2. Define the Causal Model

- Causal modeling formalizes our knowledge - however limited
  - Which variables affect each other
  - The role of unmeasured/background factors
  - The functional form of the relationships
- Focus on the structural causal model and corresponding causal graphs (Pearl2000)
  - Many other causal frameworks

2. Specify the causal model

- U: unmeasured background factors
  - e.g. stigma, partner's HIV status, ...
- W: baseline covariates
  - e.g. country, sex, age, education level, SES, ...
- A: the exposure
  - $A = 1$ for lived 1+ month outside the community
  - $A = 0$ otherwise
- Y: the outcome
  - $Y = 1$ for not testing
  - $Y = 0$ for testing

2. Specify the causal model

- The structural causal model (SCM) translates our knowledge of the study design into a set of equations
- A possible study:
  1. Randomly sample an adult
  2. Measure his/her baseline covariates
     - Region, sex, age, SES, education level, occupation ...
  3. Measure the exposure
     - "In the past year, how many months did you spend living outside the community?"
  4. Measure the outcome
     - Did the participant test at the CHC or at home?

2. Specify the causal model

The structural causal model (SCM) translates our knowledge of the study design into a set of equations

2. Specify the causal model

The structural causal model (SCM) translates our knowledge of the study design into a set of equations

Study design:

1. Sample an adult

Structural Causal Model:

$$(U_W, U_A, U_Y) \sim \mathbb{P}_U$$

2. Specify the causal model

The structural causal model (SCM) translates our knowledge of the study design into a set of equations

Study design:

1. Sample an adult
2. Measure baseline covariates

Structural Causal Model:

$$(U_W, U_A, U_Y) \sim \mathbb{P}_U$$

$$W = f_W(U_W)$$

2. Specify the causal model

The structural causal model (SCM) translates our knowledge of the study design into a set of equations

Study design:

1. Sample an adult
2. Measure baseline covariates
3. Measure the exposure (mobility)

Structural Causal Model:

$$(U_W, U_A, U_Y) \sim \mathbb{P}_U$$

$$W = f_W(U_W)$$

$$A = f_A(W, U_A)$$

2. Specify the causal model

The structural causal model (SCM) translates our knowledge of the study design into a set of equations

Study design:

1. Sample an adult
2. Measure baseline covariates
3. Measure the exposure (mobility)
4. Observe the outcome (testing)

Structural Causal Model:

$$(U_W, U_A, U_Y) \sim \mathbb{P}_U$$
$$W = f_W(U_W)$$
$$A = f_A(W, U_A)$$
$$Y = f_Y(W, A, U_Y)$$

2. Specify the causal model

The structural causal model (SCM) translates our knowledge of the study design into a set of equations

Study design:

1. Sample an adult
2. Measure baseline covariates
3. Measure the exposure (mobility)
4. Observe the outcome (testing)

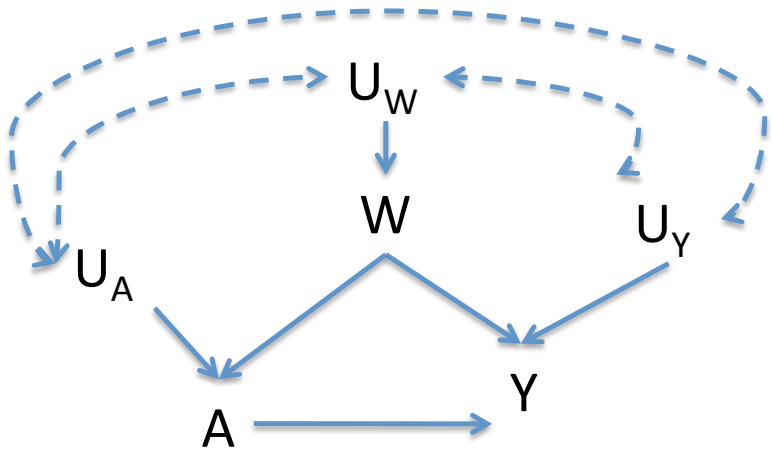
Structural Causal Model:

$$(U_W, U_A, U_Y) \sim \mathbb{P}_U$$
$$W = f_W(U_W)$$
$$A = f_A(W, U_A)$$
$$Y = f_Y(W, A, U_Y)$$

- Assumed time-ordering between variables
- No assumptions
  - On the background factors are (U_W, U_A, U_Y)
  - On the functions (f_W, f_A, f_Y)

2. Specify the causal model

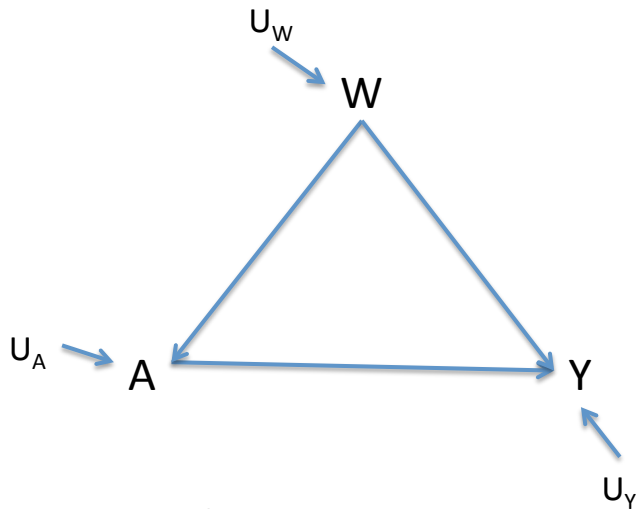
Representation as a causal graph



- The baseline covariates W represent the set of measured confounders
- The potential correlations between the unmeasured factors are represented with double-headed arrows
  - Unmeasured confounding by the shared unmeasured common causes

2. Specify the causal model

If we believed the no unmeasured confounders assumption, a possible causal graph



- Background factors are all independent
- Still no function form assumptions
- Wishing for something does not make it true

Where are we?

1. Scientific question ✓
2. Causal model ✓
3. Counterfactuals & causal parameter
4. Observed data & statistical model
5. Identifiability & statistical parameter
6. Estimation
7. Interpretation

3a. Specify the counterfactuals

- Y_1: the counterfactual testing status if, possibly contrary to fact, the adult lived 1+ month away from the community ($A = 1$)
- Y_0: the counterfactual testing status if, possibly contrary to fact, the adult lived < 1 month away from the community ($A = 0$)
- We generate counterfactuals by intervening on the causal model

$$W = f_W(U_W)$$
$$A = a$$
$$Y_a = f_Y(W, a, U_Y)$$

3b. Specify the causal parameter

Use counterfactuals to define the target causal parameter

- The difference in the expected testing uptake if all adults lived 1+ months away vs. the expected testing update if all adults lived < 1 month away:

$$\mathbb{E}[Y_1] - \mathbb{E}[Y_0]$$

  - Known as the average treatment effect (ATE)
  - For a binary outcome, the causal risk difference: $\mathbb{P}(Y_1 = 1) - \mathbb{P}(Y_0 = 1)$

- Many other causal parameters possible

3. Specify counterfactuals & the causal parameter

Why is causal inference easy for Hiro?

3. Specify counterfactuals & the causal parameter

He can time travel. He can obtain the counterfactual outcomes for all adults under the levels of the intervention of interest.

Yatta!

Where are we?

1. Scientific question ✓
2. Causal model ✓
3. Counterfactuals & causal parameter ✓
4. Observed data & statistical model
5. Identifiability & statistical parameter
6. Estimation
7. Interpretation

4a. Specify the observed data

- For one adult, the observed data are

$$O = (W, A, Y) \sim \mathbb{P}$$

  - W as measured confounders
  - A as the exposure (mobility)
  - Y as the outcome (not testing)
  - $\mathbb{P}$ as the true but unknown distribution
- In SEARCH, we have $n = 146,906$ adults with stable residence
  - We have n copies of O

4b. Link causal to observed

- We assume the causal model provides a description of our study under
  - Existing conditions (i.e. the real world)
  - Specific interventions (i.e the counterfactual world)
- This provides a link the causal world and the real (observed data) world

4c. Specify the statistical model

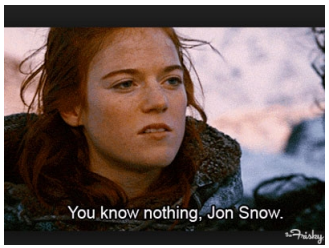
- Our causal model (what we know) $\Rightarrow$ Observed data (what we measure)
- Our causal model describes the set of processes that may have given rise to the observed data
- Our causal model implies the statistical model
  - Formally, the statistical model is the set of possible distributions of the observed data

4c. Specify the statistical model

- All statistical models are not wrong
- Our statistical model should represent real knowledge
- Causal framework helps to choose a statistical model reflecting our uncertainity
  - Often no or few restrictions on the joint distribution of the observed variables
  - e.g. Only know the exposure A is some function of baseline covariates W and unmeasured factors U_A
  - If we have real knowledge, specify it in Step 2
- Our statistical model is often non-parametric

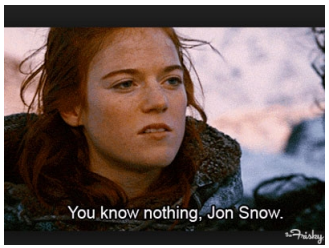
4b. Specify the statistical model

- Non-parametric: no restrictions



4b. Specify the statistical model

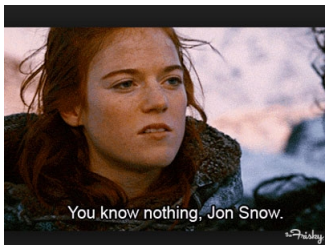
- Non-parametric: no restrictions



- Semi-parametric: some restrictions

4b. Specify the statistical model

- Non-parametric: no restrictions



- Semi-parametric: some restrictions

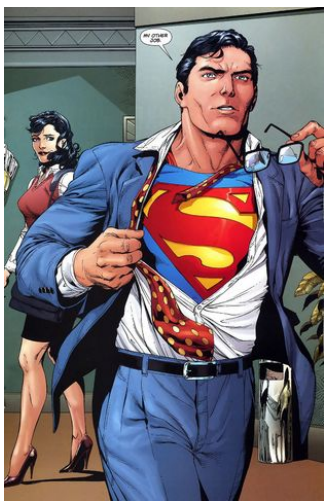
- Parametric: assumes $\mathbb{P}$ is known up to a finite number of unknown parameters

Where are we?

1. Scientific question ✓
2. Causal model ✓
3. Counterfactuals & causal parameter ✓
4. Observed data & statistical model ✓
5. Identifiability & statistical parameter
6. Estimation
7. Interpretation

5. Assess Identifiability

- Currently the parameter of interest is expressed in terms of counterfactuals: $\mathbb{E}[Y_1] - \mathbb{E}[Y_0]$
- Identifiability: what assumptions are needed to write the causal parameter as something we can estimate with the observed data?

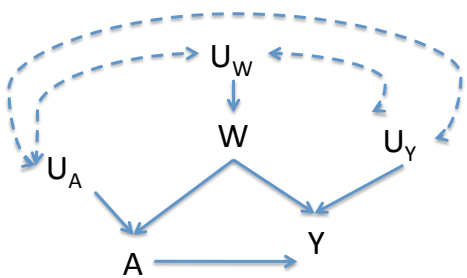


We link our day-job (estimation based on the observed data)
to our superhero-job (answering causal questions)

5. Assess Identifiability

Some intuition:

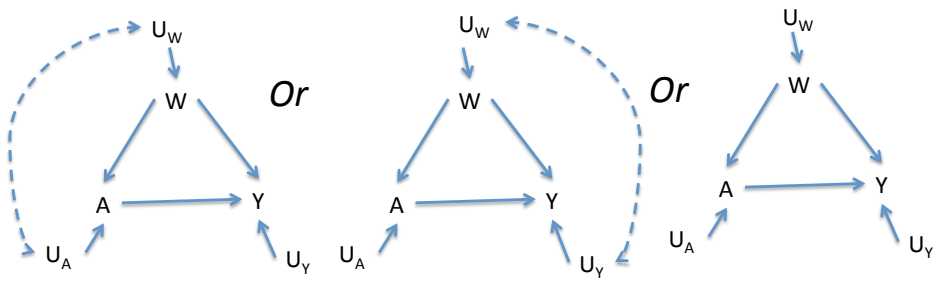
- $\mathbb{E}[Y|A=a]$: expected testing uptake among adults with mobility status $A=a$
  - Descriptive/associative
- $\mathbb{E}[Y_a]$: expected counterfactual testing uptake if all adults had mobility status $A=a$
  - Causal
- Generally $\mathbb{E}[Y|A=a]$ does *not* equal $\mathbb{E}[Y_a]$
  - Central problem in causal inference



5. Assess Identifiability

To identify our causal parameter we need:

- No unmeasured confounding
  - Equivalent to the randomization assumption: $Y_a \perp\!\!\!\perp A|W$



- Positivity: sufficient variability in the exposure within confounder strata

$$\mathbb{P}(A = a|W = w) > 0$$
$$\text{for all } w \text{ with } \mathbb{P}(W = w) > 0$$

  - Ensures the statistical parameter is well-defined

5. Assess Identifiability

With the randomization and positivity assumptions:

$$\begin{aligned}\mathbb{E}(Y_a) &= \mathbb{E}\big[\mathbb{E}(Y_a|W)\big] \\ &= \mathbb{E}\big[\mathbb{E}(Y_a|A=a,W)\big] \quad \text{under randomization} \\ &= \mathbb{E}\big[\mathbb{E}(Y|A=a,W)\big] \quad \text{under positivity}\end{aligned}$$

- Other common assumptions (temporality, stability and consistency) are implied by our causal model and the link between the causal model and statistical model
- These assumptions are not new requirements; this framework forces us to consider them explicitly

5. Assess Identifiability

The G-computation identifiability result (Robins1986):

- Under the needed assumptions:

$$\mathbb{E}(Y_1) - \mathbb{E}(Y_0) = \mathbb{E}\Big[\mathbb{E}(Y|A=1,W) - \mathbb{E}(Y|A=0,W)\Big]$$

- Difference in the expected outcome, given the exposure and confounders, and the expected outcome given no exposure and confounders, and then averaged (standardized) with respect to the covariate distribution
- For a binary outcome, equal to the marginal risk difference

$$\mathbb{E}\Big[\mathbb{P}(Y=1|A=1,W) - \mathbb{P}(Y=1|A=0,W)\Big]$$

5. Assess Identifiability

- What if the assumptions do not hold?
  - What if we do not believe the no unmeasured confounders assumption?
  - What if we do not have time-ordering?

5. Assess Identifiability

- Still have a well-defined and interpretable target parameter:
  - Difference in the marginal risk of failing to test associated with greater mobility, after controlling for the measured confounders
  - Coming as close to the wished-for causal parameter given the limitations in the data
  - More in Step 7
- Can use the lack of identifiability to inform future data collection and future studies

Where are we?

1. Scientific question ✓
2. Causal model ✓
3. Counterfactuals & causal parameter ✓
4. Observed data & statistical model ✓
5. Identifiability & statistical parameter ✓
6. Estimation
7. Interpretation

6. Estimation

- We have identified the causal parameter as a function of the observed data distribution:

$$\Psi(\mathbb{P}) = \mathbb{E}\Big[\mathbb{E}(Y|A=1,W) - \mathbb{E}(Y|A=0,W)\Big]$$

- Many estimators available:
  - Parametric G-computation (a.k.a. simple substitution estimator)
  - Inverse probability of treatment weighting (IPTW)
  - Targeted maximum likelihood estimation (TMLE)
- Nothing more-or-less causal about these estimators

6. Estimation - "Standard" approach

Pause and consider the "standard" approach

- Run logistic regression of the outcome (not testing) Y on the exposure (mobility) A and the baseline confounders W

$$logit\left[\mathbb{E}(Y|A,W)\right] = \beta_0 + \beta_1 A + \beta_2 W1 + \ldots + \beta_{19} W18$$

- Exponentiate the coefficient in front of the exposure (e^{β_1})
- Interpret as the conditional odds ratio associated with living 1+ month outside the community, while holding all the other risk factors constant

6. Estimation - "Standard" approach

Some problems:

- Our target parameter $\Psi(\mathbb{P})$ is not equal to e^{β_1}
  - Letting the estimation approach drive the question asked
  - Throwing away all our hard work!
- Relies on the main terms logistic regression being correct
  - May measure the relevant variables but do not know their exact functional relationship
  - If we had this knowledge, then we should encode it in our causal model (Step2)
  - If this parametric regression is wrong, can have biased point estimates and misleading inference

Parametric G-Computation

Consider again our target parameter:

$$\Psi(\mathbb{P}) = \mathbb{E}\big[\mathbb{E}(Y|A=1,W) - \mathbb{E}(Y|A=0,W)\big]$$
$$= \sum_w \big[\mathbb{E}(Y|A=1,W=w) - \mathbb{E}(Y|A=0,W=w)\big]\mathbb{P}(W=w)$$

1. Estimate the conditional mean outcome, given the exposure and baseline covariates $\mathbb{E}(Y|A,W)$
   - e.g. run main terms logistic regression
2. Estimate the covariate distribution $\mathbb{P}(W)$
   - Use the sample proportion $1/n \sum_{i=1}^{n} \mathbb{I}(W_i = w)$
3. Substitute in (plug-in) these estimates:

$$\Psi(\hat{\mathbb{P}}) = \frac{1}{n}\sum_{i=1}^{n}\left[\hat{\mathbb{E}}(Y_i|A_i=1,W_i) - \hat{\mathbb{E}}(Y_i|A_i=0,W_i)\right]$$

Parametric G-Computation

- Relies on consistently estimating the mean outcome $\mathbb{E}(Y|A,W)$
- Sometimes we have a lot of knowledge about the relationship between the outcome Y and the exposure-covariates (A,W)
  - If we had this knowledge, encode in our causal model and use it!
- More often, our knowledge is limited
  - Avoid introducing new assumptions during estimation
  - Assuming a parametric regression model can result in bias and misleading inferences

Inverse Probability of Treatment Weighting (IPTW)

Some Intuition:

- Can think of confounding as biased sampling
  - Certain exposure-covariate subgroups are over-represented relative to what we would see in a randomized trial
  - Other exposure-covariate subgroups are under-represented
- Apply weights to up-weight under-represented subjects and down-weight over-represented subjects
- Average and compare weighted outcomes

Estimation with IPTW

How are Inverse Probability of Treatment Weighted (IPTW) estimators like Joan from *Mad Men*?

Estimation with IPTW

Weight in all the right places

Estimation with IPTW

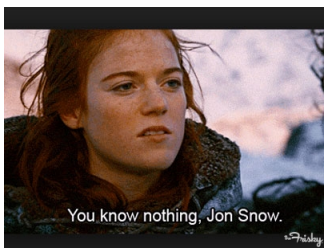
- Relies on consistently estimating the propensity score $\mathbb{P}(A = 1|W)$
- Sometimes we have a lot of knowledge about how the exposure was assigned
  - If we had this knowledge, encode in our causal model and use it!
- More often, our knowledge is limited
  - Avoid introducing new assumptions during estimation
  - Assuming a parametric regression model can result in bias and misleading inferences

Estimation with IPTW

- Tends to be an unstable estimator under positivity violations (i.e. strong confounding)
  - When covariate groups only have a few exposed or unexposed observations, weights can blow up
  - When there are covariate groups with 0 exposed or unexposed observations, weights will not blow up. BUT the estimator will likely be biased and variance underestimated
- Not guaranteed to respect the statistical model (e.g. yield probabilities less than 0 and greater than 1)
- Note: this is just one flavor of IPTW

Non-parametric Estimation

- Often our statistical model is non-parametric
- Our estimation algorithm should respect our statistical model
  - Avoid introducing new assumptions
- To estimate $\mathbb{E}(Y|A,W)$, we could take the average outcome within all strata of exposure-covariates
  - Typically have too many covariates and/or continuous covariates $\rightarrow$ empty/sparse cells
  - This approach breaks down due to the "curse of dimensionality"



Semi-parametric Estimation

- We often "know nothing", but also need to smooth over data with weak support
- Relax parametric assumptions with data-adaptive algorithms
  - e.g. stepwise regression with interactions
- However, treating the final regression as if it were pre-specified ignores the model building process
  - No reliable way to obtain inference
- Algorithm tailored to maximize/minimize some criteria and is not necessarily the best algorithm for estimating $\Psi(\mathbb{P})$

Be more flexible!

6. Estimation

We need SuperLearner!

- Flexible estimation approach to avoid unwarranted assumptions
- Uses cross-validation (sample splitting) to evaluate the performance of a library of candidate estimators

We need TMLE!

- Updates the initial estimator of $\mathbb{E}(Y|A,W)$ with information in the exposure mechanism $\mathbb{P}(A=1|W)$
  - Second chance to control for confounding
  - Hone our estimator to the parameter of interest
  - Central limit theorem for inference

Some More Notation

- $\mathbb{E}(Y|A,W)$ - the true conditional mean outcome, given the exposure and baseline covariates
- $\hat{\mathbb{E}}(Y|A,W)$ - an initial estimator based on n observations
- $\hat{\mathbb{E}}^*(Y|A,W)$ - the targeted estimator based on n observations

Overview - TMLE

1. Estimate $\mathbb{E}(Y|A,W)$ with SuperLearner
2. Estimate the propensity score $\mathbb{P}(A=1|W)$ with SuperLearner
3. Target the initial estimator $\hat{\mathbb{E}}(Y|A,W)$
4. Plug-in the updated estimates into the target parameter mapping

$$\Psi(\hat{\mathbb{P}}) = \frac{1}{n}\sum_{i=1}^{n}\left[\hat{\mathbb{E}}^*(Y_i|A_i=1,W_i) - \hat{\mathbb{E}}^*(Y_i|A_i=0,W_i)\right]$$

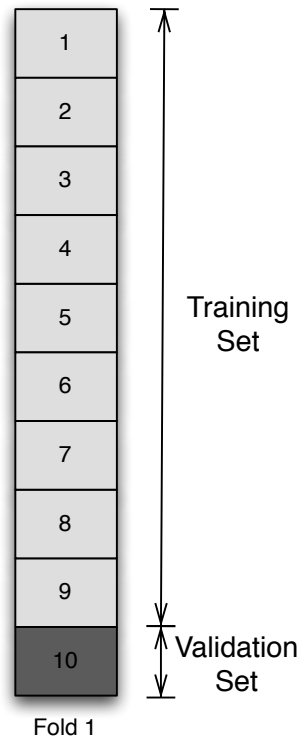
What is SuperLearner?

- Machine learning algorithm
- Uses cross-validation (data-splitting) to evaluate the performance of a library of candidate estimators
- Library can consist of a simple (e.g. main terms regression models), semi-parametric (e.g. stepwise regression, loess) and more aggressive algorithms
- Performance is measured by a loss function
  - e.g. Mean squared error (MSE)

What is SuperLearner?

Cross-validation: allows us to compare algorithms based on how they perform on independent data

- Partition the data into "folds"
- Fit each algorithm on the training set
- Evaluate its performance (called "risk") on the validation set
  - e.g. calculate the MSE for observations in the validation set
- Rotate through the folds
- Average the cross-validated risk estimates across the folds to obtain one measure of performance for each algorithm



What is SuperLearner?

- We could choose the algorithm with the best performance (i.e. smallest cross-validated risk estimate)
- Instead, SuperLearner builds the best combination of algorithm-specific estimates

Who do Captain Planet and SuperLearner need to succeed?
Our estimators combined!

Why do we need to target?

- We could use SuperLearner to predict the outcomes for all units under the treatment and control
- Then we could plug these estimates into the target parameter mapping (i.e. average the difference in the predictions):

$$\Psi(\hat{\mathbb{P}}) = \frac{1}{n}\sum_{i=1}^{n}\left[\hat{\mathbb{E}}(Y_i|A_i = 1, W_i) - \hat{\mathbb{E}}(Y_i|A_i = 0, W_i)\right]$$

- However, SuperLearner is focused on $\mathbb{E}(Y|A, W)$
  - This is not our target parameter
  - Wrong bias-variance trade-off
- Also no reliable way to obtain inference

What is targeting?

- Use information in the estimated propensity score $\hat{\mathbb{P}}(A = 1|W)$ to update the initial (SuperLearner) estimator $\hat{\mathbb{E}}(Y|A, W)$
- Involves running a univariate regression
- Use the estimated coefficient to update our initial predictions of the outcome under the treatment and under the control

Like Robin Hood, we target to hit the bullseye

How do we target?

1. Estimate the propensity score $\hat{\mathbb{P}}(A = 1|W)$
   - Again, use a flexible approach or parametric knowledge if available
2. Create the clever covariate:

$$\hat{H} = \left(\frac{\mathbb{I}(A = 1)}{\hat{\mathbb{P}}(A = 1|W)} - \frac{\mathbb{I}(A = 0)}{\hat{\mathbb{P}}(A = 0|W)} \right)$$

3. Run logistic regression of the outcome Y on the clever covariate $\hat{H}$ with offset as the logit of the initial estimates.
   - where $\text{logit}(x) = log(x/1 - x)$
4. Plug in the estimated fluctuation coefficient $\hat{\epsilon}$:

$$\text{logit}\left[\hat{\mathbb{E}}^*(Y|A, W)\right] = \text{logit}\left[\hat{\mathbb{E}}(Y|A, W)\right] + \hat{\epsilon}\hat{H}$$

TMLE - Point Estimate

5. Use the updated estimator $\hat{\mathbb{E}}^*(Y|A,W)$ to predict the outcomes for all observations under the treatment and control
6. Substitute into the target parameter mapping:

$$\Psi(\hat{\mathbb{P}}) = \frac{1}{n}\sum_{i=1}^{n}\left[\hat{\mathbb{E}}^*(Y_i|A_i=1,W_i) - \hat{\mathbb{E}}^*(Y_i|A_i=0,W_i)\right]$$

Some nice things about TMLE

- Double robust
  - Consistent if either conditional mean $\mathbb{E}(Y|A,W)$ or the propensity score $\mathbb{P}(A=1|W)$ is consistently estimated
  - Two chances!
- Semi-parametric efficient
  - Lowest asymptotic variance (most precision) among a large class if both consistently estimated
- Asymptotically linear
  - Normal curve for inference
- Substitution estimator
  - Robustness under strong confounding and rare outcomes
- Software: *tmle* and *ltmle* packages in *R*

Where are we?

1. Scientific question ✓
2. Causal model ✓
3. Counterfactuals & causal parameter ✓
4. Observed data & statistical model ✓
5. Identifiability & statistical parameter ✓
6. Estimation ✓
7. Interpretation

7. Interpretation

- Final step - consider whether and to what degree the identifiability assumptions have been met
- Statistical:
  - Estimate of the marginal difference in the risk of failing to test associated with increased mobility, after adjusting for measured confounders
  - As close as we can get to causal effect given the limitations in the data
  - "Variable importance measure"
- Causal:
  - If the necessary causal assumptions hold: Estimate of the causal risk difference or the average treatment effect

Yay!!

1. Scientific question ✓
2. Causal model ✓
3. Counterfactuals & causal parameter ✓
4. Observed data & statistical model ✓
5. Identifiability & statistical parameter ✓
6. Estimation ✓
7. Interpretation ✓

Hybrid Testing in SEARCH

A hybrid mobile approach for population-wide HIV testing in rural east Africa: an observational study

Gabriel Chamie, Tamara D Clark, Jane Kabami, Kevin Kadede, Emmanuel Ssemmondo, Rachel Steinfeld, Geoff Lavoy, Dalsone Kwarisiima, Norton Sang, Vivek Jain, Harsha Thirumurthy, Teri Liegler, Laura B Balzer, Maya L Petersen, Craig R Cohen, Elizabeth A Bukusi, Moses R Kamya, Diane V Havlir, Edwin D Charlebois

- Goal: Determine risk factors for failing to test by a hybrid testing strategy
- "Variable importance measures"
  - Determine importance of each predictor on risk of not testing, after controlling for the others

Hybrid Testing in SEARCH

- Statistical parameter - marginal relative risk:

$$\Psi(\mathbb{P}) = \frac{\mathbb{E}\left[\mathbb{E}(Y|A=1,W)\right]}{\mathbb{E}\left[\mathbb{E}(Y|A=0,W)\right]}$$

  - Each risk factor, in turn, serves as the "exposure" A and then remaining predictors as the "covariates" W
  - Estimates the marginal association after controlling for the other risk factors
  - As close to a causal interpretation given the limitations in the data
- For estimation, used TMLE with SuperLearner :)

Hybrid Testing in SEARCH

- "In multivariable analyses of adults with stable residence, predictors of non-testing included . . . migration out of the community for at least 1 month in the past year (1.60, 1.53-1.68)".
- The relative risk of not testing associated with living 1+ month away from the community was 1.60, after controlling for measured confounders
- The 95% confidence intervals were 1.53-1.68 ($p < 0.001$)

Summary & Discussion

Causal roadmap according to Jennifer Ahern

- Necessitates clearly defined scientific questions, and assures the parameters being estimated will match the questions posed
- Elaborates what assumptions are necessary to interpret an estimate causally
- When the assumptions are not met, provides guidance on how future research can be improved
- Applicable to other causal questions and data structures
  - Effects among the treated/untreated, mediation, longitudinal interventions, stochastic interventions, dynamic regimes...

Summary & Discussion

We can all be SuperLearners!!

"SuperLearner . . .
It's our hero . . .
Going to take bias down to zero"
(To the tune of "Captain Planet" theme song)

"The Power is Yours"

Summary & Discussion

TMLE as Robin Hood

- Stealing from the rich
  - Combining the best of IPTW and GComp
- and giving to the poor
  - and giving us unbiased and maximally efficient estimators

Bullseye!

A few references - not a complete bibliography

- Ahern and Balzer. Estimation and interpretation: Introduction to parametric and semi-parametric estimators for causal inference. SER Workshop, 2015.
- Hernan and Robins. Estimating causal effects from epidemiological data. *J Epidem and Community Health*, 2006.
- Pearl. *Causality: Models, Reasoning and Inference.* Cambridge University Press, 2000.
- Petersen and Balzer. Introduction to Causal Inference. UC Berkeley. 2014 `www.ucbbiostat.com`
- Petersen and van der Laan. Causal Models and Learning from Data: Integrating Causal Modeling and Statistical Estimation.*Epidemiology*, 2014.
- Robins. A new approach to causal inference in mortality studies with sustained exposure periods: application to control of the healthy worker survivor effect. *Mathematical Modelling*, 1986.
- Rosenbaum and Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 1983.
- van der Laan and Rose. *Targeted Learning: Causal Inference for Observational and Experimental Data.* Springer, 2011.

Thank you & Acknowledgements

- Jennifer Ahern
- Victor DeGruttola
- Maya Petersen
- Mark van der Laan

Supported, in part, by NIAID (R01-AI074345, U01AI099959, R37AI051164), PEPFAR, and Gilead Sciences. The SEARCH project gratefully acknowledges the Ministries of Health of Uganda and Kenya, our research team, collaborators and advisory boards, and especially all communities and participants involved.

Thank you & Questions

More info:

`http://www.ucbbiostat.com/`

lbbalzer@hsph.harvard.edu

Bonus Slides!!

5. Assess Identifiability

- Temporality: exposure precedes the outcome
  - Indicated by an arrow on the DAG from the A to Y
  - Equivalently, Y as a function of A in the causal model
- Consistency: $Y_a = Y|A = a$
  - Recall our causal model provides a description of the study under existing conditions (i.e. observed exposure) and interventions (i.e. set exposure)
- Stability: no interference between units
  - Indicated by the outcome Y being only a function of each individual's exposure A in the causal model and DAG

IPTW

More formally:

- We can re-write our target parameter as

$$\Psi(\mathbb{P}) = \mathbb{E}\left[\mathbb{E}(Y|A=1,W) - \mathbb{E}(Y|A=0,W)\right]$$
$$= \mathbb{E}\left[\left(\frac{\mathbb{I}(A=1)}{\mathbb{P}(A=1|W)} - \frac{\mathbb{I}(A=0)}{\mathbb{P}(A=0|W)}\right)Y\right]$$

  - where $\mathbb{I}(A=a)$ is an indicator function, equalling 1 if $A=a$ and 0 otherwise

- Suggests an alternate estimator:

$$\Psi(\hat{\mathbb{P}}) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\mathbb{I}(A_i=1)}{\hat{\mathbb{P}}(A_i=1|W_i)} - \frac{\mathbb{I}(A_i=0)}{\hat{\mathbb{P}}(A_i=0|W_i)}\right)Y_i$$

Step 1: Estimation with SuperLearner

Requires

- Data: $O_1, \ldots, O_n \sim \mathbb{P}_0$
- Loss function: Measure of the dissimilarity between estimate and target.
- Candidate estimators: Throw in any parametric procedure, non-parametric algorithm, histogram estimator...

Step 1: Estimation with SuperLearner

Requires

- Data: $O_1, \ldots, O_n \sim \mathbb{P}_0$
- Loss function: Measure of the dissimilarity between estimate and target.
- Candidate estimators: Throw in any parametric procedure, non-parametric algorithm, histogram estimator...

Uses Cross-Validation

- Evaluate estimator performance and prevent over-fitting

Step 1: Estimation with SuperLearner

Requires

- Data: $O_1, \ldots, O_n \sim \mathbb{P}_0$
- Loss function: Measure of the dissimilarity between estimate and target.
- Candidate estimators: Throw in any parametric procedure, non-parametric algorithm, histogram estimator...

Uses Cross-Validation

- Evaluate estimator performance and prevent over-fitting

Returns the optimal prediction function as a weighted combination of candidate estimators.

- Optimal: minimizes the expected loss, called the "risk"

How does SuperLearner work?

- Discrete super learner selects the algorithm with the smallest cross-validated risk.
- Super learner uses the predicted outcomes to create the best weighted combination of algorithms.

How does SuperLearner work?

1. Define a loss function:

$$L\big(O, \mathbb{E}(Y|A,W)\big) = \big(Y - \mathbb{E}(Y|A,W)\big)^2$$

2. Define a library of candidate estimators:

$$\mathbb{E}_{n,1}(Y|A,W) = \beta_0 + \beta_1 A + \beta_2 W_1 + \beta_3 W_2 + \beta_4 W_3$$
$$\mathbb{E}_{n,2}(Y|A,W) = \beta_0 + \beta_2 A + \beta_2 W_1 + \beta_3 sin(W_2) + \beta_4 A \; x W_1^2$$
$$\mathbb{E}_{n,3}(Y|A,W) = \text{Stepwise}$$
$$\mathbb{E}_{n,4}(Y|A,W) = \text{Loess}$$
$$\vdots$$
$$\mathbb{E}_{n,k}(Y|A,W) = \text{your advisor's favorite algorithm}$$

3. Split the data $O_1, \ldots O_n$ into $V = 10$ "folds".
   - Divide the data into ten blocks of size $n/10$.

How does SuperLearner work?

4. Define nine blocks (90% of the data) to be the training set and the remaining block (10% of the data) to be the validation set.
5. Fit each estimator on the training set.
   - e.g. Use maximum likelihood estimation to fit $\mathbb{E}_{n,1}(Y|A,W)$ on 90% of the data.
6. Predict the outcomes for the validation set.
   - e.g. Plug in the observed treatment A_i and covariates W_i for validation set (the remaining 10% of the data).

How does SuperLearner work?

7. Evaluate the empirical risk for each estimator.

$$\text{Risk}_{n,1}(v=1) = \frac{1}{n^*}\sum_{i=1}^{n^*}\left(Y_i - \mathbb{E}_{n,1}(Y_i|A_i, W_i)\right)^2$$

with n^* as the number of observations in the validation set

8. Repeat steps 4-7 so that each block gets to serve as the validation set.
9. Calculate the cross-validated risk for each algorithm.

$$\text{CV-Risk}_1 = \frac{1}{10}\sum_{v=1}^{10}\text{Risk}_{n,1}(v)$$