CDT Summer School

- Causal Excursion Effects (45 min)
- Breakout in groups + Discussion (20 min)
- Case-study of causal excursion effects with data
  - Lecture (15 min)
  - Discussion (10 min)
  - Lecture (15 min)
  - Coding in small groups (35 min)

https://github.com/StatisticalReinforcementLearningLab/Stat-ML-CDT-2023/tree/main

•Lecture 2 Causal Excursion Effects:

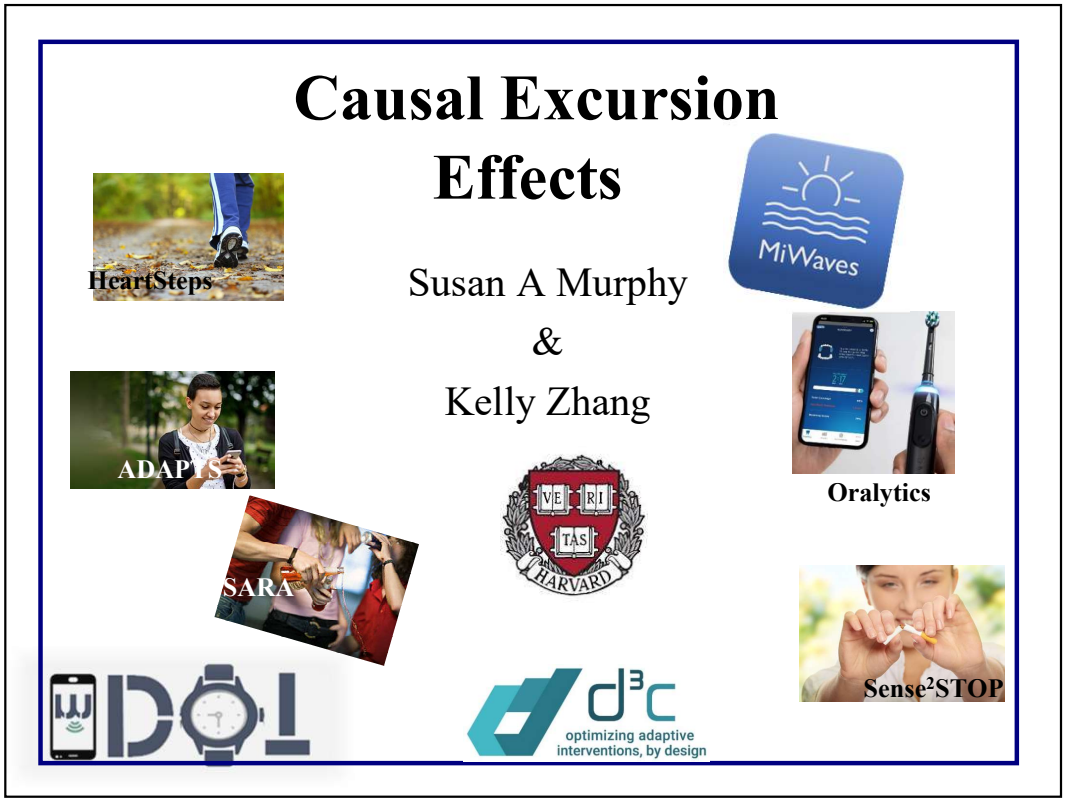
3 Hours + 15 min.13.45 to 17.00 Monday afternoon

CDT Summer School

- Break (10 min)
- Discussion: Tradeoff between within study learning (to personalize) and between study learning (generalizable knowledge)
  - Warm-start (10 min)
  - Breakout Group (10 min)
  - Discussion (10 min)

2

https://github.com/StatisticalReinforcementLearningLab/Stat-ML-CDT-2023/tree/main



some characteristics of digital health in clinical research settings

large no. of stakeholders with differing data needs

need to contribute to behavioral science

low signal to noise ratio with limited data -->intermixing behavioral science with data science

References: (see http://people.seas.harvard.edu/~samurphy/research.html)

Lecture/Practicum/Discussion 2: Analyzing Longitudinal Data Collected via MRTs and Individual-Specific RL Algorithms.

References

2.a Boruvka, A., Almirall, D., Witkiewitz, K., Murphy, S.A. (2018). Assessing Time-Varying Causal Effect Moderation in Mobile Health, Journal of the American Statistical Association. 113:523,1112-112. Accepted author version posted online: 31 Mar 2017 http://dx.doi.org/10.1080/01621459.2017.1305274. PMCID: PMC6241330

2.b Liao, P., Greenewald K., Klasnja, P. and Murphy, S.A., Personalized HeartSteps: A Reinforcement Learning Algorithm for Optimizing Physical Activity. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies March 2020 Article No.: 18. https://doi.org/10.1145/3381007 PubMed

PMCID: PMC8439432
2.c Klasnja, P., Smith, S., Seewald, N.J., Lee, A.,Hall, K., Luers, B., Hekler, E.B. and Murphy, S.A. Efficacy of contextually-tailored suggestions for physical activity: A micro-randomized optimization trial of HeartSteps Ann Behav Med. 2019 Jun; 53(6): 573–582. PubMed PMID: 30192907. PMCID: PMC6401341 Supplement

To Think About

- Why should we consider conducting simple, interpretable, primary analyses?
  - In clinical research settings
  - In commercial settings.
- What is a causal excursion effect?
- Why might we use weights to estimate a causal excursion effect?

4

HeartSteps (PI Klasnja)

Goal: Develop a mobile activity coach for individuals who are at high risk of coronary artery disease

Three iterative studies:

- 42-day micro-randomized study with sedentary individuals
- 90 day + 270 day micro-randomized, personalized, study with people who have Stage 1 Hypertension

5

https://github.com/klasnja/HeartStepsV1/wiki

Questions to Improve HeartSteps Activity Suggestions

- Do tailored activity suggestions have an effect at all?
- Do less and more burdensome activity suggestions work equally well?
- Does the effect of suggestions change over time? (e.g., do people get tired of them after a while?)
- When should we send suggestions for optimal effect?
  - Do they work better during certain parts of the day?
  - Do they work better when weather is good vs. bad?

Micro-Randomized Trial

How to justify the experimental trial costs in a clinical trials setting?

- Address a question that can be stated clearly across disciplinary boundaries and be able to provide guarantees.
- Design trial so that a variety of further interesting questions can be addressed.

First Question to Address: Do the treatments differentially impact the proximal outcome? (AKA, is there an *overall effect*?)

7

some characteristics of digital health in clinical research settings

large no. of stakeholders with differing data needs

need to contribute to behavioral science

low signal to noise ratio with limited data -->intermixing behavioral science with data science

Questions to Improve HeartSteps Activity Suggestions

- Do tailored activity suggestions have an effect at all?
- Do less and more burdensome activity suggestions work equally well?
- Does the effect of suggestions change over time? (e.g., do people get tired of them after a while?)
- When should we send suggestions for optimal effect?
  - Do they work better during certain parts of the day?
  - Do they work better when weather is good vs. bad?

MRT Data

- On each individual: $O_1, A_1, Y_2, \ldots, O_t, A_t, Y_{t+1}, \ldots$

- t: Decision point
- O_t: Observations at t^{th} decision point
- A_t: Treatment at t^{th} decision point
- Y_{t+1}: Proximal outcome

9

Notice the difference in notation as compared to this morning. Why would I use a different notation?

Recap: HeartSteps V1 MRT

- 42-day trial with 5 decision points per day
  - $t = 1, \ldots, T = 210$ decision points per user
- Observations, O_t, include recent dosage, temperature, pre-30 min step count and availability
- Treatment, A_t, is binary: whether to provide a contextually tailored activity suggestion
- Primary analysis concerns overall average effects on Y_{t+1}, 30 min step count

10

$A_t = 1$ means suggestion delivered

$I_t = 1$ means available.

Up to 210 randomizations per user

Observations: dosage, engagement indicator, temperature, location,1-hour window step count variation indicator, pre-30 step, yesterday's step count.

All decision rules/ causal inferences will be constrained by availability.

In heartsteps, decision point is available for a user if

1. user is not currently potentially operating a car, (unethical to deliver)

2. user is not currently walking, and (not scientific to deliver) another example is available only if currently classified as at risk.

3. user has not turned off intervention (user has agency)

3. user's phone phone is connected to the internet. (technical concerns) we added this when we realized there was a bug in the software code that prevented intervention delivery when phone was not connected.

Availability is not equivalent to willingness to be treated. It is momentary. Willingness to be treated/enter the study is up front. Our availability is closer to

feasibility of trt options

Adherence (i.e. compliance) is very different from availability. Suppose a decision point is available for a user at a decision point. However the phone is in their purse across the room. So they don't hear whether the phone pings/ see the lockscreen light up. This person is non-adherent at this decision point. Primary analyses will be intention-to-treat and thus will average over non-compliance.

Availability

- Interventions, A_t, can only be delivered at a decision point if the decision point is *available* for the user
  - O_t includes $I_t = 1$ if available, $I_t = 0$ if not
- Availability is known pre-decision point, i.e., pretreatment.

Availability is not the same as adherence….

11

All decision rules/ causal inferences will be constrained by availability.

In heartsteps, decision point is available for a user if

1. user is not currently potentially operating a car, (unethical to deliver)

2. user is not currently walking, and (not scientific to deliver) another example is available only if currently classified as at risk.

3. user has not turned off intervention (user has agency)

3. user's phone phone is connected to the internet. (technical concerns) we added this when we realized there was a bug in the software code that prevented intervention delivery when phone was not connected.

Availability is not equivalent to willingness to be treated. It is momentary. Willingness to be treated/enter the study is up front. Our availability is closer to feasibility of trt options

Adherence (i.e. compliance) is very different from availability. Suppose a decision point is available for a user at a decision point. However the phone is in their purse across the room. So they don't hear whether the phone pings/ see the lockscreen light up. This person is non-adherent at this decision point. Primary analyses will be intention-to-treat and thus will average over non-compliance.

Recap: MRT Data

On each of n users and at each of $t = 1, \ldots, T$ decision points:

- $A_t = 1$ if treated, $A_t = 0$ if not treated at decision t
  - Randomized,
    $$\pi_t(H_t) = P[A_t = 1 | H_t, I_t = 1]$$
- $H_t = \{O_j, A_j, Y_{j+1}\}_{j \geq 1}^{t-1} \cup \{O_t\}$

12

For simplicity, no subscript i denoting participant i.

The definition of history in this slide is different from this morning….

This morning is $H_{t-1} = \{S_j, A_j, R_{j+1}\}_{j \geq 1}^{t-1}$

Now we use $H_t = \{O_j, A_j, Y_{j+1}\}_{j \geq 1}^{t-1} \cup \{O_t\}$

H_t is the history of data for that individual. Randomization probabilities do not depend on other users' data.

Note care in subscript on Y!!!! Y_{t+1} occurs subsequent to A_t not at same time.

In heartsteps T=210 in sense2stop T=6000

Intuition

Conceptual Models

13

Conceptual Models

Generally data analysts fit a series of models:

$$Y_{t+1} \text{ “}\sim\text{” } \alpha_0 + \alpha_1^T Z_t + \beta_0 A_t$$

and then next,

$$Y_{t+1} \text{ “}\sim\text{” } \alpha_0 + \alpha_1^T Z_t + \beta_0 A_t + \beta_1 A_t X_t$$

and so on…

- Y_{t+1} is activity over 30 min. following t
- $A_t = 1$ if activity suggestion and 0 otherwise
- Z_t summaries formed from t and past/present observations
- X_t potential moderator (e.g., current weather is good or not)

14

We go through intuition that data analysts follow in clinical trial analyses in which the treatment does not vary with time. Here the complication is that treatment is time varying. The issue is that both X_t and Z_t may be outcomes of past treatment. Availability is an outcome of past treatment.

Z_t might include location, time of day, day of week, recent adherence, summaries of craving over prior hour, usual level of smoking at this time of day, etc. Might include features of time, t, so as to allow a more flexible model

X_t might be a vector as well and might include features of time X_t might be the output of a classifier or a prediction or risk formed from H_t

Above is conditional on availability

Conceptual Models

Generally data analysts fit a series of models:

Primary analysis

$$Y_{t+1} \text{ “}\sim\text{”} \alpha_0 + \alpha_1^T Z_t + \beta_0 A_t$$

and then next,

$$Y_{t+1} \text{ “}\sim\text{”} \alpha_0 + \alpha_1^T Z_t + \beta_0 A_t + \beta_1 A_t X_t$$

and so on…

Secondary analysis

- Y_{t+1} is activity over 30 min. following t
- $A_t = 1$ if activity suggestion and 0 otherwise
- Z_t summaries formed from t and past/present observations
- X_t potential moderator (e.g., current weather is good or not)

15

We go through intuition that data analysts follow in clinical trial analyses in which the treatment does not vary with time. Here the complication is that treatment is time varying. The issue is that both X_t and Z_t may be outcomes of past treatment. Availability is an outcome of past treatment.

Z_t might include location, time of day, day of week, recent adherence, summaries of craving over prior hour, usual level of smoking at this time of day, etc. Might include features of time, t, so as to allow a more flexible model

X_t might be a vector as well and might include features of time X_t might be the output of a classifier or a prediction or risk formed from H_t

Above is conditional on availability

Conceptual Models

Generally data analysts fit a series of models:

$$Y_{t+1} \text{ “}\sim\text{” } \alpha_0 + \alpha_1^T Z_t + \beta_0 A_t$$

and then next,

$$Y_{t+1} \text{ “}\sim\text{” } \alpha_0 + \alpha_1^T Z_t + \beta_0 A_t + \beta_1 A_t X_t$$

and so on…

$\alpha_1^T Z_t$ is used to reduce the noise variance in Y_{t+1}
(Z_t is sometimes called a vector of control variables)

16

We go through intuition that data analysts follow in clinical trial analyses in which the treatment does not vary with time. Here the complication is that treatment is time varying. The issue is that both X_t and Z_t may be outcomes of past treatment. Availability is an outcome of past treatment.

Z_t might include location, time of day, day of week, recent adherence, summaries of craving over prior hour, usual level of smoking at this time of day, etc. Might include features of time, t, so as to allow a more flexible model

X_t might be a vector as well and might include features of time X_t might be the output of a classifier or a prediction or risk formed from H_t

Above is conditional on availability

Causal, Marginal Effects

$$Y_{t+1} \text{ “}\sim\text{” } \alpha_0 + \alpha_1^T Z_t + \beta_0 A_t$$

β_0 is the effect, marginal over all observed and all unobserved variables, of the activity suggestion on subsequent activity.

$$Y_{t+1} \text{ “}\sim\text{” } \alpha_0 + \alpha_1^T Z_t + \beta_0 A_t + \beta_1 A_t X_t$$

$\beta_0 + \beta_1$ is the effect when the weather is good (X_t=1), marginal over other observed and all unobserved variables, of the activity suggestion on subsequent activity.

17

These are the interpretations we know hold in cross-sectional analyses and in analyses in which the treatment does not vary with time.

Goal

- Next: data analytic methods that are consistent with the scientific understanding of the meaning of the β coefficients
- Causal Inference Challenges:
  - Time-varying treatment (A_t, $t=1,\ldots T$)
  - "Independent" variables: Z_t, X_t, I_t that may be affected by prior treatment
- Robustly facilitate noise reduction via use of controls, Z_t

18

Interpretation of beta is causal and marginal

Causal Effects

- Use potential outcomes to express inferential goal mathematically
- These causal effects are *Causal Excursions*

19

Potential Outcomes

- $\bar{A}_t = \{A_1, A_2, \ldots, A_t\}$ (random variables resulting from the stochastic sampling of treatments up to and including at time t)
  - $\bar{a}_t = \{a_1, a_2, \ldots, a_t\}$ (realizations of treatments)
- $Y_{t+1}(\bar{a}_t)$ is a potential proximal outcome
- $I_t(\bar{a}_{t-1})$ is a potential "available for treatment" indicator

20

Draw attention to a_t in Y_{t+1} and a_{t-1} in I_t

Draw attention to why availability might depend on prior treatment

Nice intro to potential outcomes and causality (particularly if your data does not have randomized treatments)

Hernán MA, Robins JM (2020). Causal Inference:What If. Boca Raton: Chapman & Hall/CRC.

https://cdn1.sph.harvard.edu/wp-content/uploads/sites/1268/2020/11/ciwhatif_hernanrobins_23nov20.pdf

Potential Outcomes

- $\bar{A}_t = \{A_1, A_2, \ldots, A_t\}$
  - $\bar{a}_t = \{a_1, a_2, \ldots, a_t\}$ (realizations of treatments)
- $Y_{t+1}(\bar{a}_t)$ is a potential proximal outcome
- $I_t(\bar{a}_{t-1})$ is a potential "available for treatment" indicator
- $H_t(\bar{a}_{t-1})$ is a potential history vector
  - $X_t(\bar{a}_{t-1})$ is a vector of features of history $H_t(\bar{a}_{t-1})$

21

Note a_t in Y_{t+1} and a_{t-1} in I_t

Note that availability might depend on prior treatment

$H_t = \{O_j, A_j, Y_{j+1}\}_{j\geq 1}^{t-1} \cup \{O_t\}$; H_t is the history of data for that individual.

Nice intro to potential outcomes and causality (particularly if your data does not have randomized treatments)

Hernán MA, Robins JM (2020). Causal Inference:What If. Boca Raton: Chapman & Hall/CRC.

https://cdn1.sph.harvard.edu/wp-content/uploads/sites/1268/2020/11/ciwhatif_hernanrobins_23nov20.pdf

Fundamental Problem of Causal Inference

- Individual level causal effect (binary A_t):

$$Y_{t+1}(\bar{A}_{t-1}, 1) - Y_{t+1}(\bar{A}_{t-1}, 0)$$

$\bar{A}_{t-1} = \{A_1, A_2, \ldots, A_{t-1}\}$ (random variables resulting from the stochastic sampling of treatments up to and including at time t)

22

Nice intro to potential outcomes and causality (particularly if your data does not have randomized treatments)

Hernán MA, Robins JM (2020). Causal Inference:What If. Boca Raton: Chapman & Hall/CRC.

https://cdn1.sph.harvard.edu/wp-content/uploads/sites/1268/2020/11/ciwhatif_hernanrobins_23nov20.pdf

Fundamental Problem of Causal Inference

- Individual level causal effect (binary A_t):

$$Y_{t+1}(\bar{A}_{t-1}, 1) - Y_{t+1}(\bar{A}_{t-1}, 0)$$

- Excursion Effect

$\bar{A}_{t-1} = \{A_1, A_2, \ldots, A_{t-1}\}$ (random variables resulting from the stochastic sampling of treatments up to and including at time t)

23

Nice intro to potential outcomes and causality (particularly if your data does not have randomized treatments)

Hernán MA, Robins JM (2020). Causal Inference:What If. Boca Raton: Chapman & Hall/CRC.

https://cdn1.sph.harvard.edu/wp-content/uploads/sites/1268/2020/11/ciwhatif_hernanrobins_23nov20.pdf

Fundamental Problem of Causal Inference

- Individual level causal effect (binary A_t):

$$Y_{t+1}(\bar{A}_{t-1}, 1) - Y_{t+1}(\bar{A}_{t-1}, 0)$$

- Fundamental Problem:
  - Data alone cannot be used to predict or estimate this effect!

$\bar{A}_{t-1} = \{A_1, A_2, \ldots, A_{t-1}\}$ (random variables resulting from the stochastic sampling of treatments up to and including at time t)

24

Nice intro to potential outcomes and causality (particularly if your data does not have randomized treatments)

Hernán MA, Robins JM (2020). Causal Inference:What If. Boca Raton: Chapman & Hall/CRC.

https://cdn1.sph.harvard.edu/wp-content/uploads/sites/1268/2020/11/ciwhatif_hernanrobins_23nov20.pdf

A Solution to the Fundamental Problem of Causal Inference

Instead of trying to estimate/predict individual level causal effects,

$$Y_{t+1}(\bar{A}_{t-1}, 1) - Y_{t+1}(\bar{A}_{t-1}, 0),$$

estimate expectations or conditional expectations of these effects:

e.g.

$$E\left[Y_{t+1}(\bar{A}_{t-1}, 1) - Y_{t+1}(\bar{A}_{t-1}, 0)\right]$$

25

Nice intro to potential outcomes and causality (particularly if your data does not have randomized treatments)

Hernán MA, Robins JM (2020). Causal Inference:What If. Boca Raton: Chapman & Hall/CRC.

https://cdn1.sph.harvard.edu/wp-content/uploads/sites/1268/2020/11/ciwhatif_hernanrobins_23nov20.pdf

Marginal & Causal Effect

Excursion effect at decision point *t*:

$$E\left[Y_{t+1}\left(\bar{A}_{t-1}, 1\right) - Y_{t+1}\left(\bar{A}_{t-1}, 0\right) \mid I_t\left(\bar{A}_{t-1}\right) = 1, X_t\left(\bar{A}_{t-1}\right)\right]$$

26

Causal effect involves the randomization probabilities via the excursion.

Note conditioning on Availability

We are examining causal effects via ****excursions*** from the underlying stochastic strategy for selecting actions

X_t could be an empty set.

Marginal over randomization treatment policy (and effects thereof), conditional on availability.

The group who have the intervention turned on is unavailable and may be a select group of people-- likely depending on the intervention dose they experienced up to time t-1.

This intervention dose (all prior treatments) may have caused burden.

Marginal & Causal Effect

Excursion effect at decision point t:

$$E\left[Y_{t+1}\left(\bar{A}_{t-1}, 1\right) - Y_{t+1}\left(\bar{A}_{t-1}, 0\right) \mid I_t\left(\bar{A}_{t-1}\right) = 1, X_t\left(\bar{A}_{t-1}\right)\right]$$

– Effect is conditional on availability

27

Causal effect involves the randomization probabilities via the excursion.

Note conditioning on Availability

We are examining causal effects via ****excursions*** from the underlying stochastic strategy for selecting actions

X_t could be an empty set.

Marginal over randomization treatment policy (and effects thereof), conditional on availability.

The group who have the intervention turned on is unavailable and may be a select group of people-- likely depending on the intervention dose they experienced up to time t-1.

This intervention dose (all prior treatments) may have caused burden.

Marginal & Causal Effect

Excursion effect at decision point t:

$$E\left[Y_{t+1}\left(\bar{A}_{t-1}, 1\right) - Y_{t+1}\left(\bar{A}_{t-1}, 0\right) \mid I_t\left(\bar{A}_{t-1}\right) = 1, X_t\left(\bar{A}_{t-1}\right)\right]$$

– Effect is conditional on availability; only concerns the subpopulation of individuals available at decision t

28

Causal effect involves the randomization probabilities via the excursion.

Note conditioning on Availability

We are examining causal effects via ****excursions*** from the underlying stochastic strategy for selecting actions

X_t could be an empty set.

Marginal over randomization treatment policy (and effects thereof), conditional on availability.

The group who have the intervention turned on is unavailable and may be a select group of people-- likely depending on the intervention dose they experienced up to time t-1.

This intervention dose (all prior treatments) may have caused burden.

Marginal & Causal Effect

Excursion effect at decision point *t*:

$$E\left[Y_{t+1}\left(\bar{A}_{t-1}, 1\right) - Y_{t+1}\left(\bar{A}_{t-1}, 0\right) \mid I_t\left(\bar{A}_{t-1}\right) = 1, X_t\left(\bar{A}_{t-1}\right)\right]$$

- Effect is conditional on availability; only concerns the subpopulation of users available at decision *t*
- Effect is marginal over any $Y_u,\ u \leq t,\ A_u, u < t$ not in $X_t\left(\bar{A}_{t-1}\right)$---over all variables not in $X_t\left(\bar{A}_{t-1}\right)$.

29

Marginal & Causal Effect

Excursion effect at decision point t:

$$E\left[Y_{t+1}\left(\bar{A}_{t-1}, 1\right) - Y_{t+1}\left(\bar{A}_{t-1}, 0\right) \mid I_t\left(\bar{A}_{t-1}\right) = 1, X_t\left(\bar{A}_{t-1}\right)\right]$$

- Effect is conditional on availability; only concerns the subpopulation of users available at decision t
- Effect is marginal over any $Y_u,\ u \leq t,\ A_u, u < t$ not in $X_t\left(\bar{A}_{t-1}\right)$---over all variables not in $X_t\left(\bar{A}_{t-1}\right)$.

30

Randomized $A_t \rightarrow$

$$E\left[Y_{t+1}\left(\bar{A}_{t-1}, 1\right) - Y_{t+1}\left(\bar{A}_{t-1}, 0\right) \mid I_t\left(\bar{A}_{t-1}\right) = 1, X_t\left(\bar{A}_{t-1}\right)\right]$$

$=$

$$E\left[\ E\left[Y_{t+1} \mid A_t = 1, I_t = 1, H_t\right] - E\left[Y_{t+1} \mid A_t = 0, I_t = 1, H_t\right] \mid I_t = 1, X_t\right]$$

$=$

$$E\left[\frac{A_t Y_{t+1}}{\pi_t(H_t)} - \frac{(1 - A_t) Y_{t+1}}{\left(1 - \pi_t(H_t)\right)} \mid I_t = 1, X_t\right]$$

($\pi_t(H_t)$ is randomization probability)

31

$\pi_t(H_t) = P[A_t = 1 | H_t, I_t = 1]$

Marginal Treatment Effect

Treatment Effect Model:

$$E\left[\begin{pmatrix} E[Y_{t+1}|A_t = 1, I_t = 1, H_t] - \\ E[Y_{t+1}|A_t = 0, I_t = 1, H_t] \end{pmatrix} | I_t = 1, X_t\right] = X_t{}^T \beta$$

H_t is user's data up to and at time t
X_t is a vector of data summaries and time, t, $(X_t \subseteq H_t)$
I_t indicator of availability

We aim to conduct inference about β!

32

We are examining causal effects via ****excursions*** from the underlying stochastic strategy for selecting actions

Analysis Method: "Centered and Weighted Least Squares Estimation"

- Simple, interpretable, method for complex data
- Enables unbiased inference for a causal, overall, treatment effect (the β's)
- Inference for treatment effect is not biased by how we use the controls, Z_t, to reduce the noise variance in Y_{t+1}

references in comments! 33

Another way to think about this once you see the estimation method is to realize that now you know what you are estimating!

Boruvka, A., Almirall, D., Witkiewitz, K., Murphy, S.A. (2018). Assessing Time-Varying Causal Effect Moderation in Mobile Health, Journal of the American Statistical Association. 113:523,1112-112. Accepted author version posted online: 31 Mar 2017 http://dx.doi.org/10.1080/01621459.2017.1305274. PMCID: PMC6241330

Qian, T., Yoo, H., Klasnja, P., Almirall, D. and Murphy, S.A., (2021) Estimating Time-Varying Causal Excursion Effects in Mobile Health with Binary Outcomes with discussion. *Biometrika* Volume 108, Issue 3, September 2021, Pages 507–527. DOI 10.1093/biomet/asaa070. Rejoinder. *Biometrika* Volume 108, Issue 3, September 2021, Pages 551–555

Estimation

- Select probabilities: $\tilde{p}_t(s) \in (0,1)$
- Form weights: $W_t = \left(\frac{\tilde{p}_t(X_t)}{\pi_t(H_t)}\right)^{A_t} \left(\frac{1-\tilde{p}_t(X_t)}{1-\pi_t(H_t)}\right)^{1-A_t}$
- Center treatment actions: $A_t \rightarrow \left(A_t - \tilde{p}_t(X_t)\right)$
- Minimize:

$$E_n\left[\sum_{t=1}^{T}\left(Y_{t+1} - Z_t{}^T\alpha - \left(A_t - \tilde{p}_t(X_t)\right)X_t{}^T\beta\right)^2 I_t W_t\right]$$

- E_n is empirical distribution over individuals.

34

$Z_t{}^T\alpha$ is working model for mean of Y_{t+1} given I_t=1 and H_t

This model is not assumed to be correct. It is used for variance reduction.

Minimize

$$E_n\left[\sum_{t=1}^{T}\left(Y_{t+1}-Z_t{}^T\alpha\ -\left(A_t-\tilde{p}_t(X_t)\right)X_t{}^T\beta\right)^2 I_t W_t\right]$$

Good but ***incorrect*** intuition:

Weighted least squares with a centered treatment indicator, $A_t - \tilde{p}_t(S_t)$ thus using the assumption:

- $E[Y_{t+1}|A_t, I_t = 1, Z_t] = Z_t{}^T\alpha + \left(A_t - \tilde{p}_t(X_t)\right)X_t{}^T\beta$

E_n is expectation with respect to empirical distribution

This intuition is not correct because we are not assuming that the conditional mean of Y_{t+1} given Z_t, X_t, I_t=1, has the form $Z_t{}^T\alpha$!

Minimize

$$E_n\left[\sum_{t=1}^{T}\left(Y_{t+1}-Z_t{}^T\alpha\ -\left(A_t-\tilde{p}_t(X_t)\right)X_t{}^T\beta\right)^2 I_tW_t\right]$$

Good but ***incorrect*** intuition:

- $E[Y_{t+1}|A_t, I_t=1, Z_t] \neq Z_t{}^T\alpha + \left(A_t-\tilde{p}_t(X_t)\right)X_t{}^T\beta$

36

Minimize

$$E_n\left[\sum_{t=1}^{T}\left(Y_{t+1}-Z_t{}^T\alpha-\left(A_t-\tilde{p}_t(X_t)\right)X_t{}^T\beta\right)^2 I_t W_t\right]$$

The Modeling Assumption:

$$E\left[\left(\begin{array}{c}E[Y_{t+1}|A_t=1,I_t=1,H_t]-\\ E[Y_{t+1}|A_t=0,I_t=1,H_t]\end{array}\right)|I_t=1,X_t\right]=X_t{}^T\beta_0$$

If $\tilde{p}_t$ depends at most on X_t, then, under moment conditions, $\hat{\beta}$ is consistent for β_0

37

If in the W_t you included [\tilde p_t(X_t) (1-\tilde p_t(X_t))]^{-1} then you would be obtaining the "Best Linear Predictor" of the treatment effect. (we recommend doing this in future…)

Theory

Under moment conditions, the large sample distribution of $\sqrt{n}(\hat{\beta} - \beta_0)$ is approximately a Normal distribution with mean 0 and var-covar matrix, $(\sum_p)^{-1} \sum (\sum_p)^{-1}$

$$\sum_p = \mathrm{E}\left[\sum_{t=1}^{T} \tilde{p}_t(X_t)\left(1 - \tilde{p}_t(X_t)\right) I_t X_t X_t^T\right]$$

Z_t and X_t are finite dimensional feature vectors.

38

See https://arxiv.org/abs/2203.13887 for more sophisticated use of action-centering ("Neyman orthogonality)

Gains from Randomization

- Causal inference for a marginal treatment effect
- Inference on treatment effect is robust to working model:

$$E[Y_{t+1} | I_t = 1, H_t] \approx Z_t^T \alpha$$

  - $Z_t \subseteq H_t$
  - Contrast to literature on partially linear, single index models and varying coefficient models

39

Choice of Weights

Choice of $\tilde{p}_t(X_t)$ determines marginalization over time under model misspecification of treatment effect.

Example: $X_t = 1$, $\tilde{p}_t(X_t)$=$\tilde{p}$. Resulting $\hat{\beta}$ is an estimator of

$$\sum_{t=1}^{T} E[I_t]\beta_t \Big/ \sum_{t=1}^{T} E[I_t]$$

where

$$\beta_t = E[\ E[Y_{t+1}|A_t = 1, I_t = 1, H_t] - E[Y_{t+1}|A_t = 0, I_t = 1, H_t]|\ I_t = 1\ ]$$

40

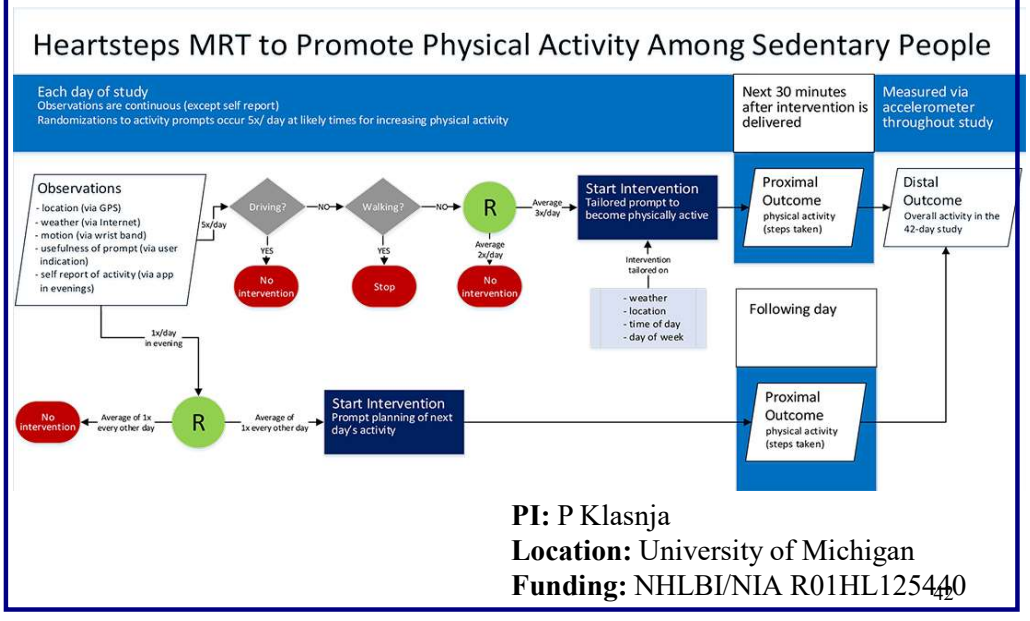
Interpretation when modeling assumption does not hold.

HeartSteps V1

41

https://github.com/klasnja/HeartStepsV1/wiki

HeartSteps V1



This project tests the feasibility and effectiveness of providing, via a smartphone, just-in-time tailored physical activity suggestions as well as evening prompts to plan the following day's physical activity so as to help sedentary individuals increase their activity. The resulting data will be used to inform the development of a JITAI for increasing physical activity. 42 days; n=37

PI: Predrag Klasnja

Location: University of Michigan

Funding: NHLBI/NIA R01HL125440

heartsteps MRT
https://www.clinicaltrials.gov/ct2/show/NCT03225521?titles=HeartSteps&rank=1

Klasnja, P., Smith, S., Seewald, N.J., Lee, A.,Hall, K., Luers, B., Hekler, E.B. and Murphy, S.A. Efficacy of contextually-tailored suggestions for physical activity: A micro-randomized optimization trial of HeartSteps Ann Behav Med. 2019 Jun; 53(6): 573–582. HeartSteps V1 data PubMed PMID: 30192907. PMCID: PMC6401341

On each of $n=37$ participants:

a) Activity suggestion, A_t

- **Provide a suggestion with probability .6**
  - a tailored sedentary-reducing activity suggestion (probability=.3)
  - a tailored walking activity suggestion (probability=.3)
- **Do nothing (probability=.4)**

- 5 times per day * 42 days= 210 decision points

43

Conceptual Models

$$Y_{t+1} \text{ “}\sim\text{”} \alpha_0 + \alpha_1 Z_t + \beta_0 A_t$$
$$Y_{t+1} \text{ “}\sim\text{”} \alpha_0 + \alpha_1 Z_t + \alpha_2 d_t + \beta_0 A_t + \beta_1 A_t d_t$$

- $t=1,\ldots T=210$
- Y_{t+1} = log-transformed step count in the 30 minutes *after* the t^{th} decision point,
- $A_t = 1$ if an activity suggestion is delivered at the t^{th} decision point; $A_t = 0$, otherwise,
- Z_t = log-transformed step count in the 30 minutes *prior* to the t^{th} decision point,
- d_t =days in study; takes values in (0,1,….,41)

44

9.1 Interactions with A

App use on prior day

Discounted number of messages sent in past

Other location indicator (1 if not at work or home)

Std dev. of steps in 60 min window in previous 7 days.

9.2 Control

Above interaction features

Log-transformed tracker steps 30 mins prior to decision point

sq. root number of steps yesterday

Temperature

HeartSteps V1 Analysis

$$Y_{t+1} \text{ “}\sim\text{”} \ \alpha_0 + \alpha_1 Z_t + \beta_0 A_t, \text{ and}$$

$$Y_{t+1} \text{ “}\sim\text{”} \ \alpha_0 + \alpha_1 Z_t + \alpha_2 d_t + \beta_0 A_t + \beta_1 A_t d_t$$

Causal Effect Term	Estimate	95% CI	p-value
$\beta_0 A_t$ *(effect of an activity suggestion)*	$\hat{\beta}_0$=.13	(-0.01, 0.27)	.06
$\beta_0 A_t + \beta_1 A_t d_t$ *(time trend in effect of an activity suggestion)*	$\hat{\beta}_0$ = .51	(.20, .81)	<.01
	$\hat{\beta}_1$ = -.02	(-.03, -.01)	<.01

45

.13 translates into a 14% increase over no treatment in step count about 33 steps

mean 30-minute step count is 253 steps

.51 translates into a 67% increase over no treatment in step count about 170 steps

Midway through study d_t=20 this increase has reduced to 16% increase in step count

On each of $n=37$ participants:

a) Activity suggestion

- Provide a suggestion with probability .6
  - **a tailored walking activity suggestion (probability=.3)**
  - **a tailored sedentary-reducing activity suggestion (probability=.3)**
- **Do nothing (probability=.4)**

- 5 times per day * 42 days= 210 decision points

46

HeartSteps Analysis

$$Y_{t+1} \text{ “}\sim\text{” } \alpha_0 + \alpha_1 Z_t + \beta_0 A_{1t} + \beta_1 A_{2t}$$

- $A_{1t} = 1$ if walking activity suggestion is delivered at the t^{th} decision point; $A_{1t} = 0$, otherwise,
- $A_{2t} = 1$ if sedentary-reducing activity suggestion is delivered at the t^{th} decision point; $A_{2t} = 0$, otherwise,

Causal Effect	Estimate	95% CI	p-value
$\beta_0 A_{1t} + \beta_1 A_{2t}$	$\hat{\beta}_0 = .21$ $\hat{\beta}_1 > 0$	(.04, .39) ns	.02 ns

47

mean 30-minute step count is 253 steps

.21 translates into a 23% increase over no treatment in step count about 59 steps

When d_t is added to the model one finds that initially beta0 coefficient of A1t=.729 and the coefficient of A1t*d_t is -.025

Pvalues for both are .000

.729 translates into a 107% increase over no treatment in step count about 271 steps

Initial Conclusions

- The data indicates that there is a causal effect of the activity suggestion on step count in the succeeding 30 minutes.
  - This effect is primarily due to the walking activity suggestions.
  - This effect deteriorates with time
  - The walking activity suggestion initially increases step count over succeeding 30 minutes by approximately 271 steps but by day 20 this increase is only approximately 65 steps.

48

Discussion

Problematic Analyses

- GLM & GEE analyses
- Random effects models & analyses
- Machine Learning Generalizations:
  - Partially linear, single index models & analysis
  - Varying coefficient models & analysis

--These analyses do not take advantage of the micro-randomization. Can accidentally eliminate the advantages of randomization for estimating causal effects--

49

In terms of the ability to obtain causal marginal effects and in terms of robustness.

SEMIPARAMETRIC GEE ANALYSIS IN PARTIALLY LINEAR SINGLE-INDEX MODELS FOR LONGITUDINAL DATA

BY JIA CHEN, DEGUI LI, HUA LIANG†,1 AND SUOJIN WANG‡,2 2015, Vol. 43, No. 4, 1682–1715

New Estimation and Model Selection Procedures for Semiparametric Modeling in Longitudinal Data Analysis, Jianqing FAN and Runze LI Journal of the American Statistical Association

September 2004, Vol. 99,pg. 710

References

Intensive Longitudinal Methods by Niall Bolger and Jean-Philippe Laurenceau (2013)

Models for Intensive Longitudinal Data edited by Walls and Schaer (2006)

A Time-Varying Effect Model for Intensive Longitudinal Data by Tan et al., *Psychol Methods*. 2012 March ; 17(1): 61–77.

This last paper does not have the problem as long as there are no subject specific

random effects.

Dynamical systems analyses, e.g time series or pomps or mps

20 min for Discussion!

Break & Discussion

- What exactly was the benefit of randomization?
- What would you do to the weighted least squares criterion to obtain an estimator of the BLP: Best Linear Predictor?
- Hard: Can you include in the weighted least squares criterion an adjustment for correlated errors?

51

extra slides

Price due to Marginal Estimand

This "least squares-like" method does not allow for working non-independence correlation matrix

– Estimating function is biased if off-diagonal elements in working correlation matrix: in general,

$$E\Big[\big(Y_{t+1} - Z_t{}^T\alpha - \big(A_t - \tilde{p}_t(X_t)\big)X_t{}^T\beta\big)I_t W_t I_u W_u \sigma_{t,u} Z_u\Big] \neq 0$$

if $u \neq t$

53

Such a result is unsurprising given the bias that arises under

non-independence structures in IPTW (Vansteelandt 2007; Tchetgen Tchetgen et al. 2012) or

in GEEs where a time-varying response is modelled by time-varying covariates (Pepe and

Anderson 1994; Schildcrout and Heagerty 2005).