Assessing Time-Varying Causal Interactions and Treatment Effects with Applications to Mobile Health

Susan Murphy

01.11.17

MD2K
Center of Excellence for Mobile Sensor Data-to-Knowledge

The Methodology Center
advancing methods, improving health

MICHIGAN STATISTICS

INSTITUTE FOR SOCIAL RESEARCH
UNIVERSITY OF MICHIGAN

Outline

- Introduction to mobile health
- Causal Treatment Effects (aka Causal Excursions)
- (A wonderfully simple) Estimation Method
- HeartSteps

HeartSteps

Context provided via data from:

Wearable band → activity and sleep quality;

Smartphone sensors → busyness of calendar, location, weather;

Self-report → stress, user burden

In which contexts should the smartphone provide the user with an activity suggestion?

Data from wearable devices that sense and provide treatments

- 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 (high dimensional)
- A_t: Action at t^{th} decision point (treatment)
- Y_{t+1}: Proximal response (e.g., reward, utility, cost)

Examples

1) Decision Points (Times, *t*, at which a treatment might be provided.)
   1) Regular intervals in time (e.g. every minute)
   2) At user demand

Heart Steps: approximately every 2-2.5 hours: pre-morning commute, mid-day, mid-afternoon, evening commute, after dinner

Examples

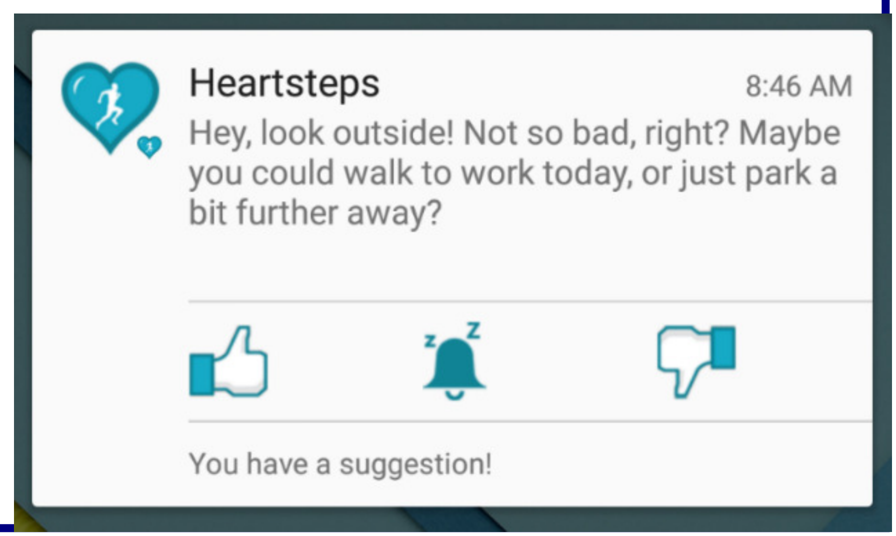
2) Observations O_t
   1) Passively collected (via sensors)
   2) Actively collected (via self-report)

<u>Heart Steps</u>: classifications of activity, location, weather, step count, busyness of calendar, user burden, adherence…….

Examples

3) Actions A_t
    1) Types of treatments that can be provided at a decision point, t
    2) Whether to provide a treatment

<u>HeartSteps</u>: tailored activity suggestion (yes/no)



Availability

- Treatments, A_t, can only be delivered at a decision point if an individual is *available*.
  - O_t includes I_t=1 if available, I_t=0 if not

- Treatment effects at a decision point are conditional on availability.

- Availability is not the same as adherence!

Examples

4) Proximal Response Y_{t+1}

Heart Steps: Step count over next 30 minutes (activity suggestions)

Continually Learning Mobile Health Intervention

1) Trial Designs: Are there effects of the actions on the proximal response? *experimental design*

2) Data Analytics for use with trial data: Do effects vary by the user's internal/external context,? Are there delayed effects of the actions? *causal inference*

3) Learning Algorithms for use with trial data: Construct a "warm-start" treatment policy. *batch Reinforcement Learning*

4) Online Algorithms that personalize and continually update the mHealth Intervention. *online Reinforcement Learning*

Micro-Randomized Trial Data

On each of n participants and at each of T decision points:

- O_t observations at decision point t,
  - includes I_t=1 if available, I_t=0 if not
- A_t=1 if treated, A_t=0 if not treated at decision t
  - Randomized, $P[A_t=1 | H_t, I_t=1]=p_t(H_t)$
- Y_{t+1} proximal response

$H_t=\{(O_i, A_i, Y_{i+1}), i=1,\ldots,t-1; O_t\}$ denotes data through t

Conceptual Models

Generally data analysts fit a series of increasingly more complex models:

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

and then next,

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

and so on…

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

Conceptual Models

Generally data analysts fit a series of increasingly more complex models:

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

and then next,

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

and so on…

$\alpha_0 + \alpha_1^T Z_t$ is used to reduce the noise variance in Y_{t+1}

(Z_t is sometimes called a vector of control variables)

Causal 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.

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

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

Our Goal

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

Outline

- Introduction to mobile health
- Causal Treatment Effects (aka Causal Excursions)
- (A wonderfully simple) Estimation Method
- HeartSteps

Potential Outcomes

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

Marginal & Causal Effect

Effect at decision point *t*:

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

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

Marginal & Causal Effect

Lagged effect at decision point *t*:

$$E\left[Y_{t+2}\left(\bar{A}_{t-1}, 1, A_{t+1}^{a_t=1}\right) - Y_{t+2}\left(\bar{A}_{t-1}, 0, A_{t+1}^{a_t=0}\right) \mid I_t(\bar{A}_{t-1}) = 1, S_t(\bar{A}_{t-1})\right]$$

- Impact of $A_{t+1}^{a_t=1}$ depends on randomization probabilities
- Effect is marginal over any $Y_u,\ u \leq t,\ A_u, u < t$ not in $S_t(\bar{A}_{t-1})$----over all variables not in $S_t(\bar{A}_{t-1})$.
- Effect is conditional on availability; only concerns the subpopulation of available individuals.
- Definitions for greater lags are similar.

Consistency & Micro-Randomized $A_t \rightarrow$

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

$=$

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

$=$

$$E\left[\frac{A_t Y_{t+1}}{p_t(H_t)} - \frac{(1 - A_t) Y_{t+1}}{\left(1 - p_t(H_t)\right)}\ | I_t = 1, S_t\right]$$

($p_t(H_t)$ is randomization probability)

Marginal Treatment Effect

Treatment Effect Model:

$$E[\ E[Y_{t+1}|A_t = 1, I_t = 1, H_t] \\ - E[Y_{t+1}|A_t = 0, I_t = 1, H_t]|\ I_t = 1, S_t] \\ = {S_t}^T \beta$$

H_t is participant's data up to and at time t

S_t is a vector of data summaries and time, t, ($S_t \subseteq H_t$)

I_t indicator of availability

We aim to conduct inference about β!

Outline

- Introduction to mobile health
- Causal Treatment Effects (aka Causal Excursions)
- (A wonderfully simple) Estimation Method
- HeartSteps

“Centered and Weighted Least Squares Estimation”

- Simple method for complex data
- Enables unbiased inference for a causal, marginal, 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}

https://arxiv.org/abs/1601.00237

Estimation

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

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

- E_n is empirical distribution over individuals.

Minimize

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

- Good but incorrect intuition: Weighted "GEE" with a working independence correlation matrix and a centered treatment indicator, $A_t - \tilde{p}_t(S_t)$

E_n is expectation with respect to empirical distribution

Minimize

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

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

Model assumption:

$$E[(E[Y_{t+1}|A_t = 1, I_t = 1, H_t] - E[Y_{t+1}|A_t = 0, I_t = 1, H_t])|I_t = 1, S_t] = {S_t}^T\beta_0$$

Theory

Under moment conditions, $\sqrt{n}(\hat{\beta} - \beta_0)$ converges to a Normal distribution with mean *0* and var-covar matrix, $\left(\sum_p\right)^{-1} \Sigma \left(\sum_p\right)^{-1}$

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

Z_t and S_t are finite dimensional feature vectors.

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

Price due to Marginal Estimand

This "GEE-like" method can only use a working independence correlation matrix

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

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

if $u \neq t$

Choice of Weights

Choice of $\tilde{p}_t(S_t)$ determines marginalization over time.

Example: $S_t = 1$, $\tilde{p}_t(S_t)$=$\tilde{p}$. 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

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

Outline

- Introduction to mobile health
- Causal Treatment Effects (aka Causal Excursions)
- (A wonderfully simple) Estimation Method
- HeartSteps

Heart Steps Pilot Study

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

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)

Pilot Study 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 $\hat{\beta}_1$ = -.02	(.20, .81) (-.03, -.01)	<.01 <.01

Heart Steps Pilot Study

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

Pilot Study 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

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 ≈ 271 steps but by day 21 this increase is only ≈ 65 steps.

Discussion

Problematic Analyses

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

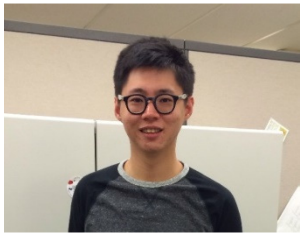
--These analyses do not take advantage of the micro-randomization. Can accidentally eliminate the advantages of randomization for estimating causal effects--

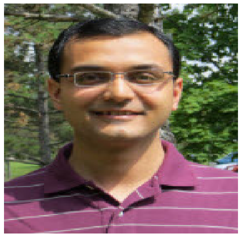
Discussion

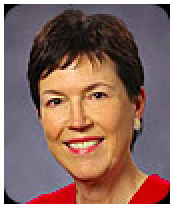
- Randomization enhances:
  - Causal inference based on minimal structural assumptions

- Challenge:
  - How to include random effects which reflect scientific understanding ("person-specific" effects) yet not destroy causal inference?

Collaborators!







mHealth

Sense2Stop Smoking Cessation Coach

- Wearable wrist/chest bands provide multiple physiological sensor streams…; craving, burden,…..
- Supportive stress-regulation interventions available on smartphone 24/7
- In which contexts should the wrist band provide reminder to access stress-regulation apps?

Minimize

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

- Resulting $\hat{\beta}$ is an estimator of

$$\left(E\sum_{t=1}^{T}\tilde{p}_t(S_t)\big(1-\tilde{p}_t(S_t)\big)I_t S_t S_t{}^T\right)^{-1} E\left[\sum_{t=1}^{T} S_t\tilde{p}_t(S_t)\big(1-\tilde{p}_t(S_t)\big)I_t\beta_t(S_t)\right]$$