# Causal Excursion Effects (Individual RL Algorithms)

Monday Afternoon Session
Kelly Zhang and Susan Murphy

# Just-In-Time Adaptive Interventions (JITAIs)

- Interventions that are delivered **whenever and wherever needed**

- Examples

  - Heartsteps

  - Oralytics

  - Sense2Stop

- **Goal of Micro-Randomized Trial:** To inform the design of JITAIs

# Questions to Inform the Design of JITAI
## (from Susan's Heartsteps Slides)

- 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?
- Is the suggestion effectiveness, including states in which they work, different for different types of people?

# What is a Micro-Randomized Trial?

(1) Intervening on people repeatedly over an extended period of time

(2) Treatments or Actions $A_{i,t}$ are randomized.

$\boxed{\text{States } S_{i,t} \text{ is a subset of } O_{i,t}}$

For each user $i \in [1:n]$,

$\boxed{\text{Time-Varying Covariates } X_{i,t} \text{ is a subset of } O_t}$

$$\underbrace{\left(O_{i,1}, A_{i,1}, Y_{i,2}\right)}_{D_{i,1}} \quad \underbrace{\left(O_{i,2}, A_{i,2}, Y_{i,3}\right)}_{D_{i,2}} \quad \dots \quad \underbrace{\left(O_{i,T}, A_{i,T}, Y_{i,T+1}\right)}_{D_{i,T}}$$

# How to probabilistically choose actions?

**(1) Randomize with a constant probability**

- $A_{i,t} \sim \mathrm{Bernoulli}(p)$

**(2) Randomize with a constant probability when person is available**

- HeartSteps V1:  $p = 0.6$  when not driving, walking, etc. otherwise, $p = 0$

  - Cannot answer any causal questions about "unavailable" times

- **Drawback:** User burden and habituation due to interrupting people in states for which they are not responsive.

# How to probabilistically choose actions?

**(1) Ra**

- *A*

<div style="border:2px solid black; background:#e8e8e8; padding:10px;">

**Data Collection Policy –>**

**What questions you can assess after study is over**

</div>

**(2) Randomize with a constant probability when person is available**

- HeartSteps V1: $p = 0.6$ when not driving, walking, etc. otherwise, $p = 0$

  - Cannot answer any causal questions about "unavailable" times

- **Drawback:** User burden and habituation due to interrupting people in states for which they are not responsive.

# How to probabilistically choose actions?

**(3) Randomization Depends on Person's State**

- Interventions may only be useful in certain states

- **Example:** Mornings randomize with $p = 0.8$ and evenings randomize with $p = 0.2$

- Can be informed by prior data

- **Drawbacks:**

  - Prior data not always available, or prior data from target population may not be available

  - Distribution shift and non-stationarity between prior data and MRT

# How to probabilistically choose actions?

## (4) Stochastic Online Algorithm

- Rather than choosing a policy apriori, specify some objective that the decision making algorithm optimizes online

- **Example:** Use posterior sampling RL algorithm to bias the randomization in favor of actions that should maximize rewards (Heartsteps V2/V3)

- **Example:** Heartsteps anti-sedentary messages - spread treatments uniformly across anti-sedentary times

# Suppose we ran an MRT…

**Do the treatments differentially impact the proximal outcome?**

- In certain states?

- For people with certain traits?

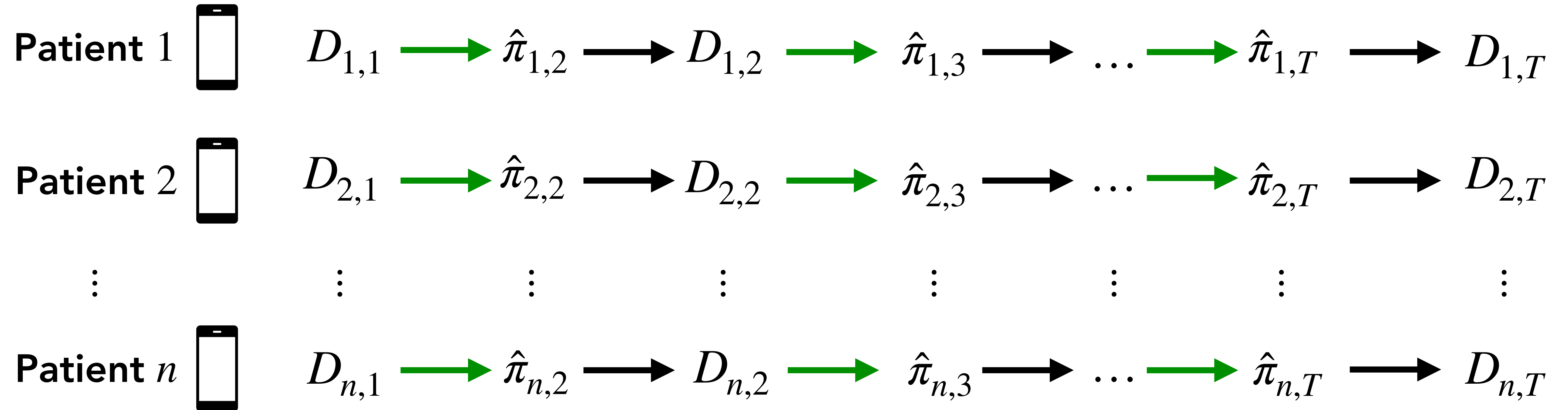- On average?

**Standard Treatment Effect**

$Y(0)$  $Y(1)$

$$\mathbb{E}\left[Y(1) - Y(0) \mid X = x\right]$$

$$D_{i,t} \triangleq (O_{i,t}, A_{i,t}, R_{i,t+1})$$

# Longitudinal Patient Data

Algorithm Update

Data Collection

**Patient** $1$ $\quad D_{1,1} \longrightarrow \hat{\pi}_{1,2} \longrightarrow D_{1,2} \longrightarrow \hat{\pi}_{1,3} \longrightarrow \dots \longrightarrow \hat{\pi}_{1,T} \longrightarrow D_{1,T}$

**Patient** $2$ $\quad D_{2,1} \longrightarrow \hat{\pi}_{2,2} \longrightarrow D_{2,2} \longrightarrow \hat{\pi}_{2,3} \longrightarrow \dots \longrightarrow \hat{\pi}_{2,T} \longrightarrow D_{2,T}$

**Patient** $n$ $\quad D_{n,1} \longrightarrow \hat{\pi}_{n,2} \longrightarrow D_{n,2} \longrightarrow \hat{\pi}_{n,3} \longrightarrow \dots \longrightarrow \hat{\pi}_{n,T} \longrightarrow D_{n,T}$

Outcomes are dependent over time within for each patient

# Issues in the MRT setting…

- Delayed effects of treatments

    - Responsiveness to treatment today depends on previous treatments

- Non-Stationarity
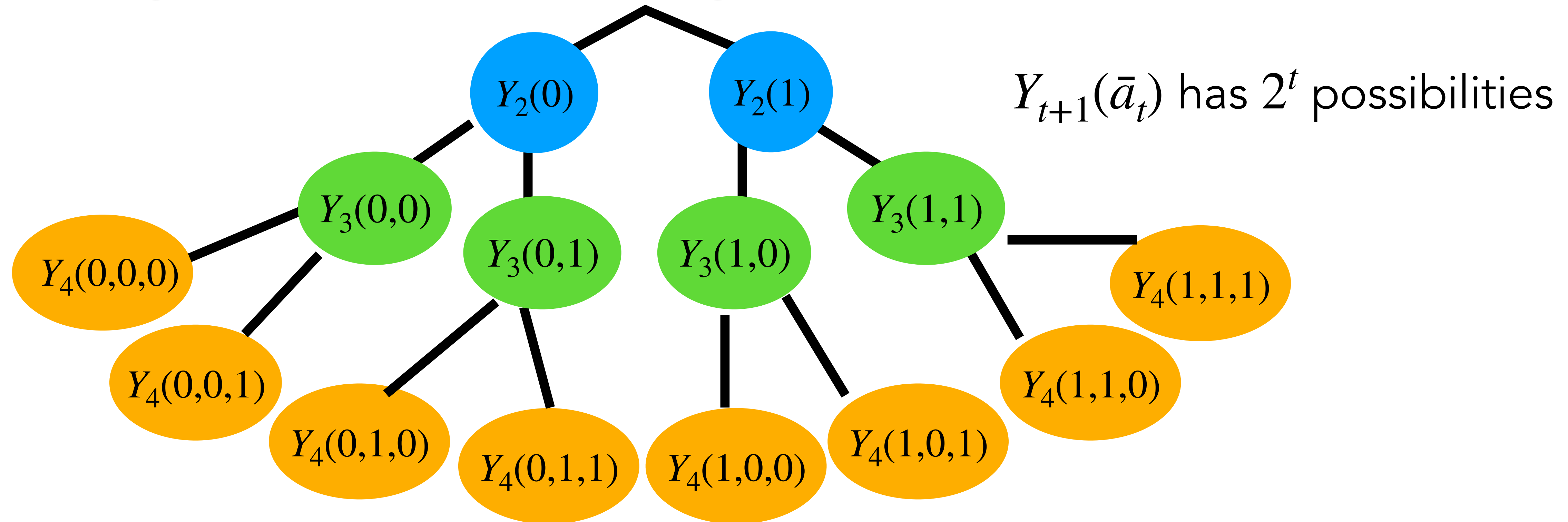
    - Responsiveness to treatment changes over time

**Not Captured by "Standard Treatment Effect"**

$Y(0)$   $Y(1)$

$$\mathbb{E}\left[Y(1) - Y(0) \mid X = x\right]$$

# Longitudinal Data Setting: Potential Outcomes



$Y_{t+1}(\bar{a}_t)$ has $2^t$ possibilities

(1) Patients (potential outcome tree) drawn i.i.d. from a population

(2) Patients "tree" of potential outcomes:

$$\left\{ O_{i,t}(\bar{a}_{t-1}), Y_{i,t+1}(\bar{a}_t) : \bar{a}_t \in \{0,1\}^t \right\}_{t=1}^{T}$$

# Longitudinal Data Setting: Treatment Effects

**Do the treatments differentially impact the proximal outcome on average given time-varying covariates $x$?**

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

where $\bar{A}_{t-1} = \{A_1, A_2, \ldots, A_{t-1}\}$

Averages over randomness in

(1) Draw of patient from population (potential outcomes)

(2) Randomness in previous action selection $\bar{A}_{t-1}$, aka "behavior policy"

# Interpretation as Causal "Excursion" Effects

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

where $\bar{A}_{t-1} = \{A_1, A_2, \ldots, A_{t-1}\}$

- Above represents the effect of treating vs not treating given time-varying covariates $x$

  - At time $t$, when the behavior policy is used to select previous actions $\bar{A}_{t-1}$

- Represents the effect of taking an "one time-step excursion" from the behavior policy

# Inferential Goal: Causal Excursion Effect

$$\mathbb{E}\left[Y_{t+1}(\bar{A}_{t-1},1) - Y_{t+1}(\bar{A}_{t-1},0) \,|\, X_t = x\right] = f(x)^{\top}\theta$$

where $\bar{A}_{t-1} = \{A_1, A_2, \ldots, A_{t-1}\}$ and $f(x)$ is a feature mapping

We are interested in the best fitting linear model, i.e., some $\theta^{\star}$

# Small Group Discussion Questions

(1) Why might the state $S_t$ used by an RL algorithm might be different from time-varying covariates $X_t$ used in conditional excursion effects?

(2) Recall the model $\mathbb{E}\left[Y_{t+1}(\bar{A}_{t-1},1) - Y_{t+1}(\bar{A}_{t-1},0)\,|\,X_t = x\right] = f(x)^\top\theta$. The HeartSteps RL algorithm had a different model of the rewards / outcomes. How is this coherent? **What is the purpose of having an after study analyses model that is separate RL algorithm model?**

(3) Are there other causal excursion effects that one might be interested in besides $\mathbb{E}\left[Y_{t+1}(\bar{A}_{t-1},1) - Y_{t+1}(\bar{A}_{t-1},0)\,|\,X_t = x\right]$?

# We are given the following MRT dataset…

Simple behavior policy

- If participant has been recently physically active, randomize with $p = 0.3$

- Otherwise, randomize with $p = 0.5$

**Task:** We suppose the following model for the excursion effect

$$\mathbb{E}\left[Y_{t+1}(\bar{A}_{t-1},1) - Y_{t+1}(\bar{A}_{t-1},0) \,|\, X_t = x\right] = f(x)^{\top}\theta$$

How can we estimate some best fitting $\theta$?

# Form a model for $\mathbb{E}\left[Y_{t+1} \mid H_{t-1}, X_t\right]$

We can form a model for $Y_{t+1}$

$$\mathbb{E}\left[Y_{t+1}(\bar{A}_{t-1}, A_t) \mid H_{t-1}, X_t, A_t\right] = g(H_{t-1}, X_t)^\top \eta + A_t f(X_t)^\top \theta$$

- Model for excursion effect

$$\mathbb{E}\left[Y_{t+1}(\bar{A}_{t-1}, 1) - Y_{t+1}(\bar{A}_{t-1}, 0) \mid H_{t-1}, X_t\right] = f(X_t)^\top \theta$$

- Model for baseline outcome

$$\mathbb{E}\left[Y_{t+1}(\bar{A}_{t-1}, 0) \mid H_{t-1}, X_t\right] = g(H_{t-1}, X_t)^\top \eta$$

# Fit model for $\mathbb{E}\left[Y_{t+1} \mid H_{t-1}, X_t\right]$

**Least Squares Loss:**

$$\ell(H_{i,t}; \eta, \theta) \triangleq \sum_{t=1}^{T} \left(Y_{i,t+1} - g(H_{i,t-1}, X_{i,t})^{\top}\eta - A_{i,t}f(X_{i,t})^{\top}\theta\right)^2$$

**Loss Minimizer:**

$$(\hat{\eta}, \hat{\theta}) = \operatorname{argmin}_{\eta,\theta} \frac{1}{n} \sum_{i=1}^{n} \ell(H_{i,t}; \eta, \theta)$$

**Equivalently $(\eta, \theta) = (\hat{\eta}, \hat{\theta})$ solves:**

$$0 = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial(\eta, \theta)} \ell(H_{i,t}; \eta, \theta)\Big|_{(\eta,\theta)}$$

Fit model for $\mathbb{E}\left[Y_{t+1} \mid H_{t-1}, X_t\right]$

**Least Squares Loss:**

$$\ell(H_{i,t}; \eta, \theta) \triangleq \sum_{t=1}^{T} \left(Y_{i,t+1} - g(H_{i,t-1}, X_{i,t})^{\top}\eta - A_{i,t} f(X_{i,t})^{\top}\theta\right)^2$$

**When do we expect the above approach to work well vs. not?**

- **Concern:** When the model for outcome under $A_t = 0$ is poorly specified $\mathbb{E}\left[Y_{t+1}(\bar{A}_{t-1}, 0) \mid H_{t-1}, X_t\right] = g(H_{t-1}, X_t)^{\top}\eta$

  - $\eta$ is a nuisance parameter

- **Concern:** Excursion effect model $\mathbb{E}\left[Y_{t+1}(\bar{A}_{t-1}, 1) - Y_{t+1}(\bar{A}_{t-1}, 0) \mid H_{t-1}, X_t\right]$ depends other entries in $H_{t-1}$ (besides $X_t$)

# Estimation with Nuisance Parameter

**Ideal Scenario:** Estimate $\theta$ consistently even if our model for the baseline outcome $\mathbb{E}\left[Y_{t+1}(\bar{A}_{t-1}, 0) \mid H_{t-1}, X_t\right] = g(H_{t-1}, X_t)^\top \eta$ is wrong

**Turns out this is possible!**

**High-Level Idea:** Rewrite the estimation criteria so that misspecification of the baseline model does not affect the estimation of $\theta$ in the limit (Neyman orthogonalization)

# Action Centering

**Least Squares Loss:**

$$\ell(H_{i,t}; \eta, \theta) \triangleq \sum_{t=1}^{T} \left( Y_{i,t+1} - g(H_{i,t-1}, X_{i,t})^\top \eta - (A_{i,t} - \pi_{i,t})f(X_{i,t})^\top \theta \right)^2$$

**Why is this loss reasonable for estimating $\theta$?**

- $\mathbb{E}\left[ Y_{t+1}(\bar{A}_{t-1}, A_t) \mid H_{t-1}, X_t \right] = g(H_{t-1}, X_t)^\top \eta$ now model of **average** reward

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

**Critical Assumption:** $\pi_{i,t}$ depends only on $X_{i,t}$

# Why is adding action centering helpful?

$$\frac{\partial}{\partial \theta} \ell(H_{i,t}; \eta, \theta) \triangleq \sum_{t=1}^{T} \left\{ Y_{i,t+1} - g(H_{i,t-1}, X_{i,t})^{\top} \eta - (A_{i,t} - \pi_{i,t}) f(X_{i,t})^{\top} \theta \right\} (A_{i,t} - \pi_{i,t}) f(X_{i,t})$$

- $\theta = \theta^{\star}$ solves $0 = \mathbb{E}\left[\dfrac{\partial}{\partial \theta} \ell(H_{i,t}; \eta, \theta)\right]$

  - The solution $\theta^{\star}$ is not affected by $\eta$ at all!!!

- Key observation:

$$\mathbb{E}\left[g(H_{i,t-1}, X_{i,t})^{\top} \eta \left(A_{i,t} - \pi_{i,t}\right) f(X_{i,t}) \middle| H_{i,t-1}, X_{i,t}\right]$$

$$= g(H_{i,t-1}, X_{i,t})^{\top} \eta \mathbb{E}\left[A_{i,t} - \pi_{i,t} \middle| H_{i,t-1}, X_{i,t}\right] f(X_{i,t}) = 0$$

# Why is adding action centering helpful?

**Why include $g(H_{i,t-1}, X_{i,t})^\top \eta$ in the model at all?**

- Action centering ensures the limiting $\theta^\star$ is not affected by the model $g(H_{i,t-1}, X_{i,t})^\top \eta$

- In small samples $\hat{\theta}$ could have lower variance if $g(H_{i,t-1}, X_{i,t})^\top \eta$ helps reduce noise

- Can replace $g(H_{i,t-1}, X_{i,t})^\top \eta$ with neural network / random forest using double machine learning (large data setting) [Chernozhukov, 2018]

# What if the MRT data was collected with an RL algorithm?

- Complication is that $\pi_{i,t}$ no longer depends on just $X_{i,t}$, but also $H_{i,t-1}$
- Need to incorporate weights to ensure that action centering trick "works"
  - i.e., only need to assume that the excursion effect model is correct $\mathbb{E}\left[Y_{t+1}(\bar{A}_{t-1}, 1) - Y_{t+1}(\bar{A}_{t-1}, 0) \mid X_t\right] = f(X_t)^\top \theta$

- Weights $W_{i,t} = \left(\dfrac{p(X_{i,t})}{\pi_{i,t}}\right)^{A_{i,t}} \left(\dfrac{1 - p(X_{i,t})}{1 - \pi_{i,t}}\right)^{1 - A_{i,t}}$

See "*Note on Excursion Effects and Action Centering*" for formal justification

$$(\hat{\eta}, \hat{\theta}) = \operatorname{argmin}_{\eta, \theta} \frac{1}{n} \sum_{i=1}^{n} \sum_{t=1}^{T} W_{i,t} \left\{ Y_{i,t+1} - g(H_{i,t-1}, X_{i,t})^\top \eta - (A_{i,t} - p(X_{i,t})) f(X_{i,t})^\top \theta \right\}^2$$