

# Online RL Algorithms that Pool Across Users

Tuesday Morning Session  
Kelly Zhang and Susan Murphy

# Collaborators!



Susan Murphy



Lucas Janson



Anna Trella



Inbal Nahum-Shani



Vivek Shetty

# Digital Intervention Study Design Objectives

## Within-Study Personalization

Maximize User Benefit

- Send messages at opportune moments

## Use Online RL Algorithms

$$\mathbb{E} \left[ \sum_{t=1}^T R_t \right]$$

## After-Study Analyses

Evaluate the Intervention

- Understand heterogeneity across user types and user states

Infer Treatment Effects

$$\mathbb{E}[Y_{t+1}(\bar{A}_{t-1}, 1) - Y_{t+1}(\bar{A}_{t-1}, 0) | X_t]$$

# Digital Intervention Study Design Objectives

## Within-Study Personalization

Maximize User Benefit

- Send messages at opportune moments

## Use Online RL Algorithms

$$\mathbb{E} \left[ \sum_{t=1}^T R_t \right]$$

## After-Study Analyses

**Confidence Intervals Critical for**

- Replicable science
- Publishing and sharing results

Infer Treatment Effects

$$\mathbb{E}[Y_{t+1}(\bar{A}_{t-1}, 1) - Y_{t+1}(\bar{A}_{t-1}, 0) | X_t]$$

## To think about!

- How do we balance these objectives of within-study personalization and after-study analyses?
- How does using RL algorithms that learn across users make after study analyses more challenging?

# Oralytics Study Overview

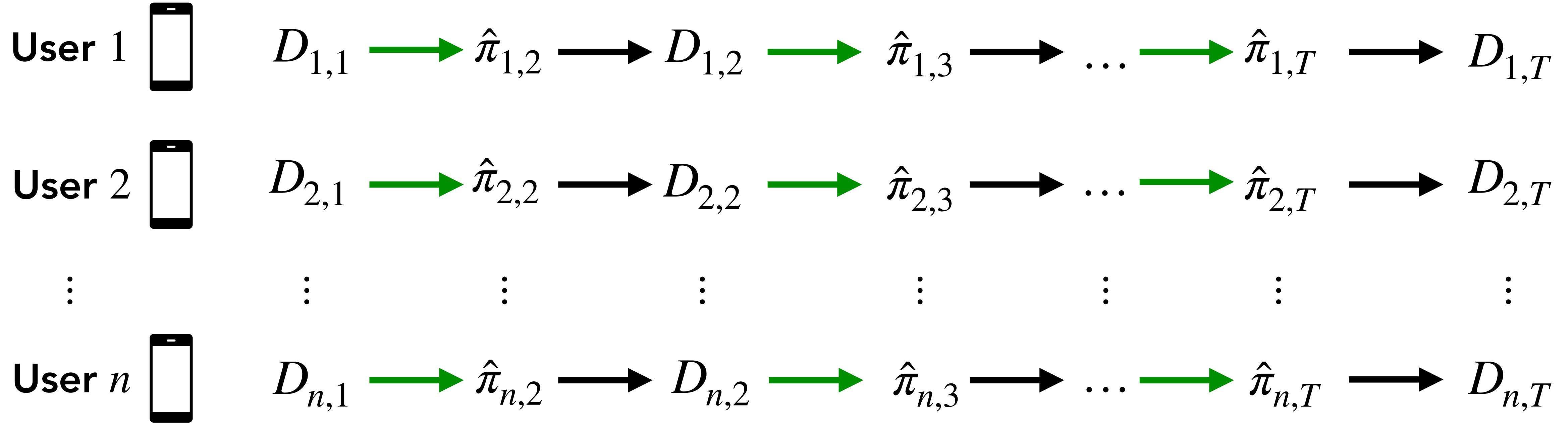
- **Total Decision Times:** 10 weeks with two decision times per day ( $T = 140 = 10 \cdot 7 \cdot 2$ )
- **Study Population:**  $N \approx 70$  patients from dental clinics in Los Angeles
- **Data Collected After Study:** For each user  $i \in [1 : N]$ ,

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

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

## Individual RL Algorithms

→ Algorithm Update  
→ Data Collection



### Dependence Within a User

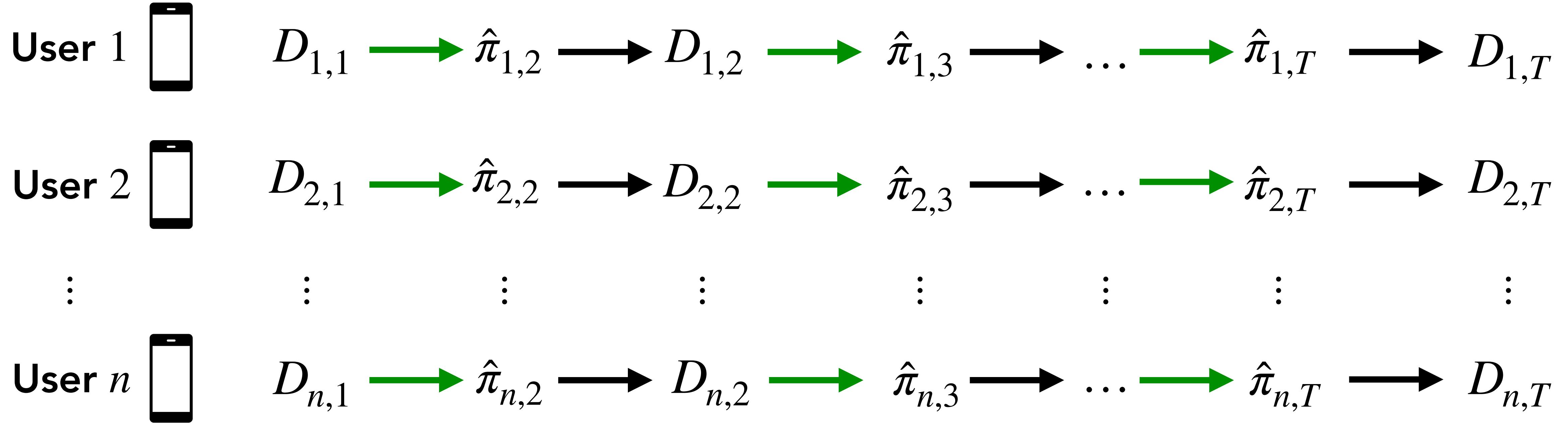
User states/rewards can be dependent over time

### Limitations?

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

## Individual RL Algorithms

→ Algorithm Update  
→ Data Collection



### Dependence Within a User

User states/rewards can be dependent over time

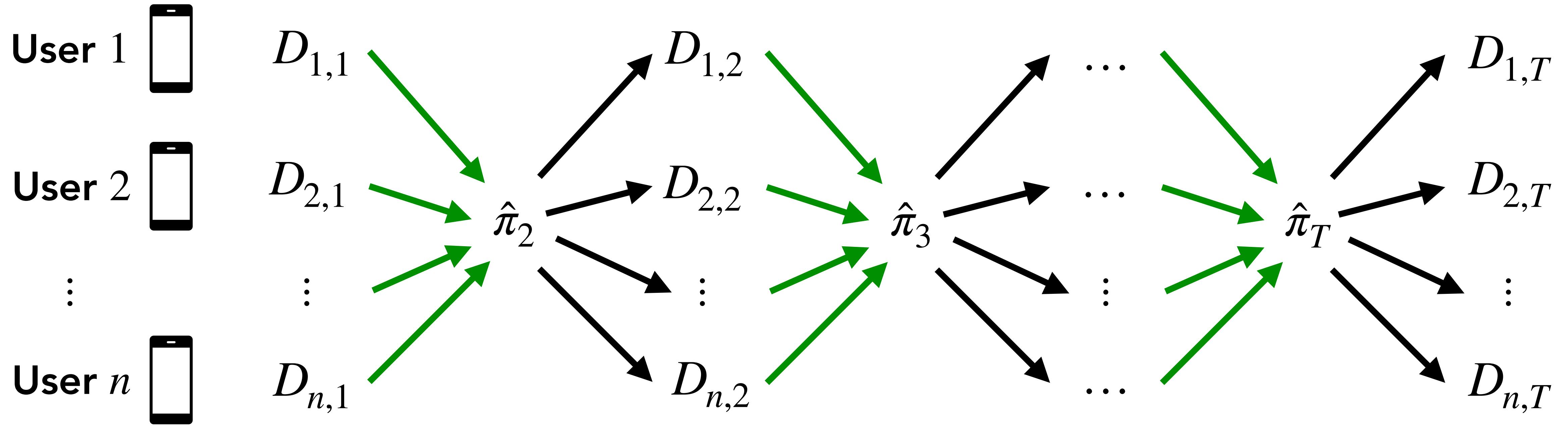
### Limitations

Rewards are noisy and few decision times per user → slow learning

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

# Pooling RL Algorithm

→ Algorithm Update  
→ Data Collection



## Dependence Within a User

User states/rewards can be dependent over time

## Dependence Between Users

Due to use of pooling algorithm

# Related Work

## Inference after Adaptive Sampling

- However, assume contextual bandit environment
- Hadad et al., 2021; Zhang et al. 2021; Bibaut et al. 2021

## Inference for Longitudinal Data

- Does not allow for pooled RL algorithms (assumes i.i.d. user trajectories)
- Boruvka et al. 2016; Qian et al. 2020

## This work

- Inference after using pooled RL algorithms for longitudinal data environments
- Make stronger assumptions on the pooled RL algorithm since we control it and have full knowledge of it

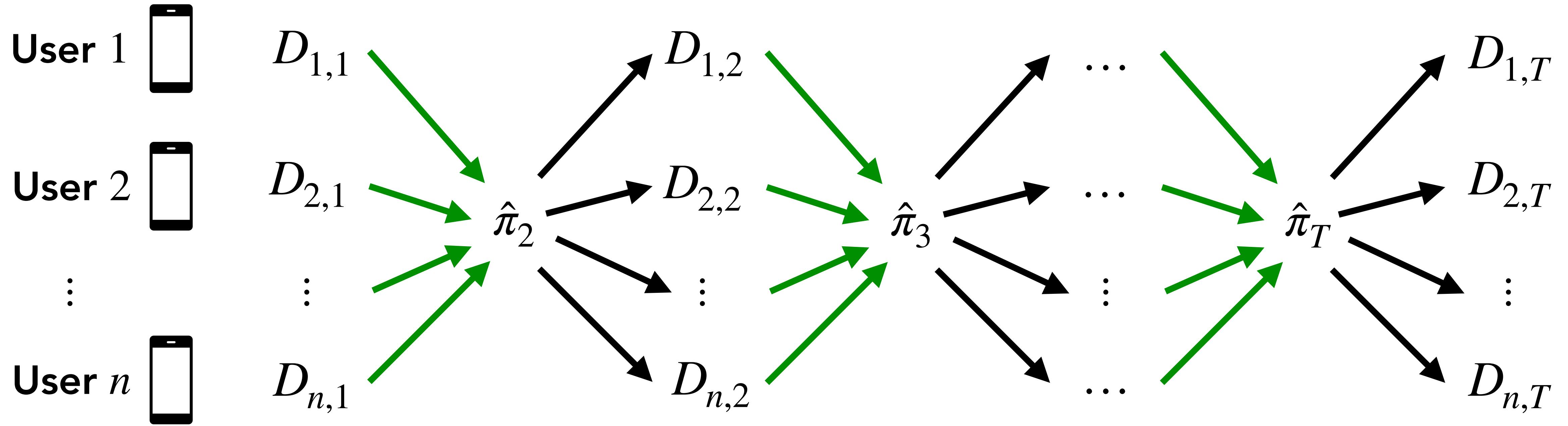
# Overview

- 1. Excursion effects after pooling**
2. Overview of Inferential Approach
3. Asymptotic Normality Proof Ideas

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

# Pooling RL Algorithm

→ Algorithm Update  
→ Data Collection



## Dependence Within a User

User states/rewards can be dependent over time

## Dependence Between Users

Due to use of pooling algorithm

# Excursion Effects under Pooling

Recall the excursion effects we considered with no pooling:

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

- Randomness is over (i) potential outcomes, and (ii)  $\bar{A}_{t-1}$ , aka “behavior policy”

**Why is the above problematic as  $n$  grows when there is pooling?**

- Under pooling, the distribution of  $\bar{A}_{t-1}$  depends on how many other users are in the study!!

# What can we do?

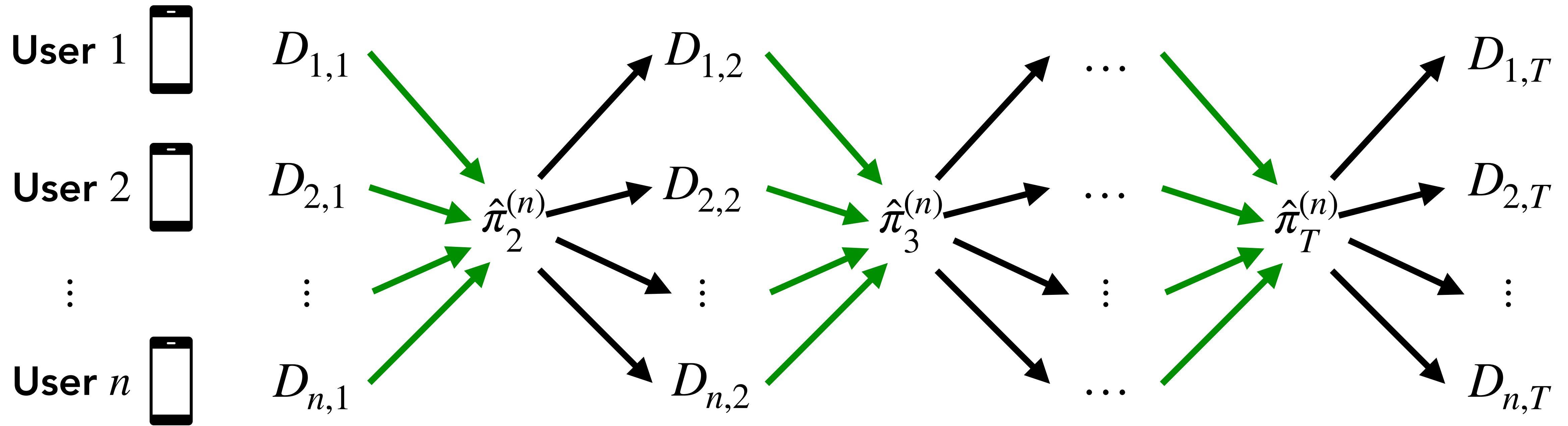
**Idea:** If the pooling policy converges to a limiting policy as  $n \rightarrow \infty$ , then we can consider excursions from that limiting policy

## Excursion Effect from Limiting Policy:

$$\mathbb{E}_{\pi^*} [Y_{t+1}(\bar{A}_{t-1}, 1) - Y_{t+1}(\bar{A}_{t-1}, 0) | X_t = x]$$

- Randomness is over (i) potential outcomes, and (ii)  $\bar{A}_{t-1}$  chosen according to the limiting policy  $\pi^*$

# Key Assumption: Convergence to Limiting Policies



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

For each  $\hat{\pi}_t^{(n)}$  as  $n \rightarrow \infty$ ,  
 $\hat{\pi}_t^{(n)} \rightarrow \pi_t^*$  (limiting policy)

**Limiting policy:** the policy that would be learned if deployed on the whole population

# Assumption: Parametric Policy Classes

**Policy Class:**  $\{\pi(\cdot; \beta)\}_{\beta \in \mathbb{R}^d}$

- Estimated policy:  $\hat{\pi}_t^{(n)}(s) \triangleq \pi(s; \hat{\beta}_{t-1}^{(n)})$
- Limiting policy:  $\pi_t^\star(s) \triangleq \pi(s; \beta_{t-1}^\star)$

Form  $\hat{\beta}_{t-1}^{(n)}$  with  $\{H_{i,t-1}\}_{i=1}^n$   
(e.g. estimate of reward  
model parameters)

# Assumption: Parametric Policy Classes

**Policy Class:**  $\{\pi(\cdot; \beta)\}_{\beta \in \mathbb{R}^d}$

- Estimated policy:  $\hat{\pi}_t^{(n)}(s) \triangleq \pi(s; \hat{\beta}_{t-1}^{(n)})$
- Limiting policy:  $\pi_t^\star(s) \triangleq \pi(s; \beta_{t-1}^\star)$

Form  $\hat{\beta}_{t-1}^{(n)}$  with  $\{H_{i,t-1}\}_{i=1}^n$   
(e.g. estimate of reward  
model parameters)

## Key Assumptions

1. Convergence of  $\hat{\beta}_t^{(n)} \xrightarrow{P} \beta_t^\star$  (for each  $t$ )
2. Policy class  $\{\pi(\cdot; \beta)\}_{\beta \in \mathbb{R}^d}$  is smooth in  $\beta$  (Lipschitz)

# Assumption: Parametric Policy Classes

## Example RL Algorithm: Boltzmann Sampling

$$\begin{aligned} P(A_{i,t+1} = 1 | H_{1:n,t}, S_{i,t+1}) &= \text{sigmoid}(\phi(S_{i,t+1})^\top \hat{\beta}_t) \\ &= \frac{1}{1 + \exp(-\phi(S_{i,t+1})^\top \hat{\beta}_t)} \end{aligned}$$

## Key Assumptions

1. Convergence of  $\hat{\beta}_t^{(n)} \xrightarrow{P} \beta_t^*$  (for each  $t$ )
2. Policy class  $\{\pi(\cdot; \beta)\}_{\beta \in \mathbb{R}^d}$  is smooth in  $\beta$  (Lipschitz)

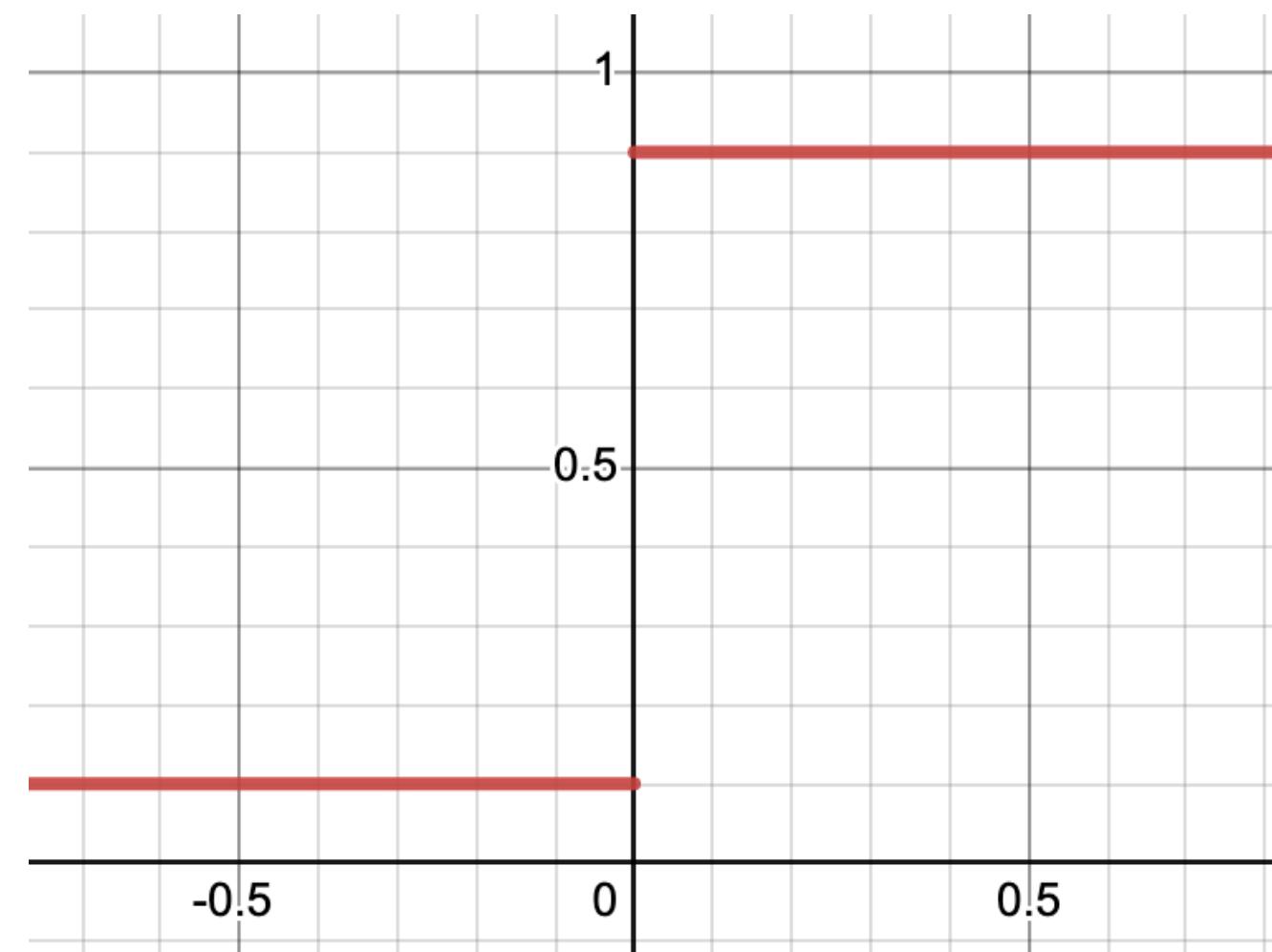
No assumption that RL algorithm's model is correct

# Allocation Function: What probability should the limiting policy send a message?

**Maximize Rewards**

$$\pi^*(s) = \mathbf{1}\{\text{Treatment Effect}(s) > 0\}$$

Probability  
of Sending a  
Message

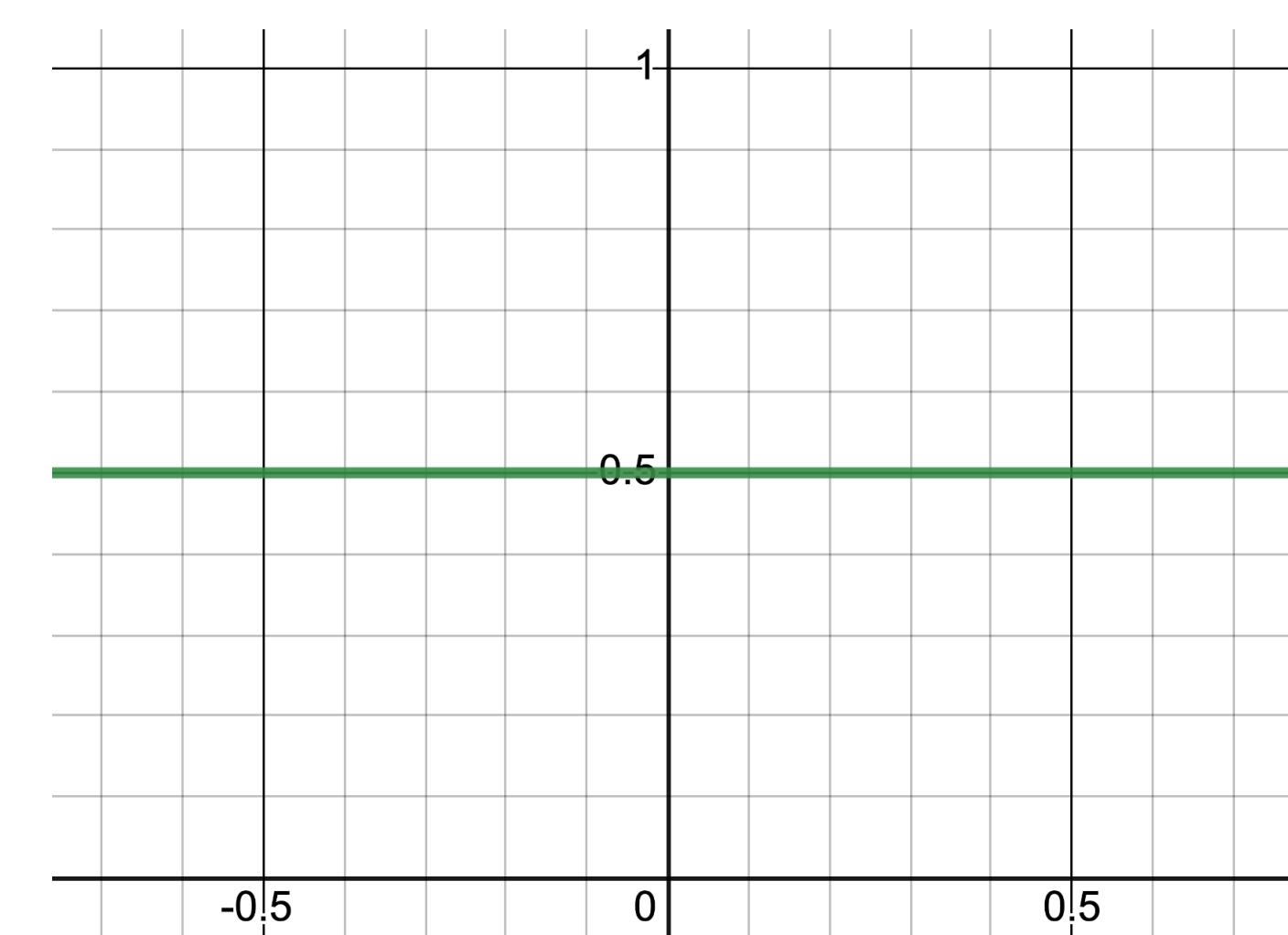


Treatment Effect in State  $s$

**Accurately Infer Treatment Effects**

$$\pi^*(s) = 0.5$$

Probability  
of Sending a  
Message



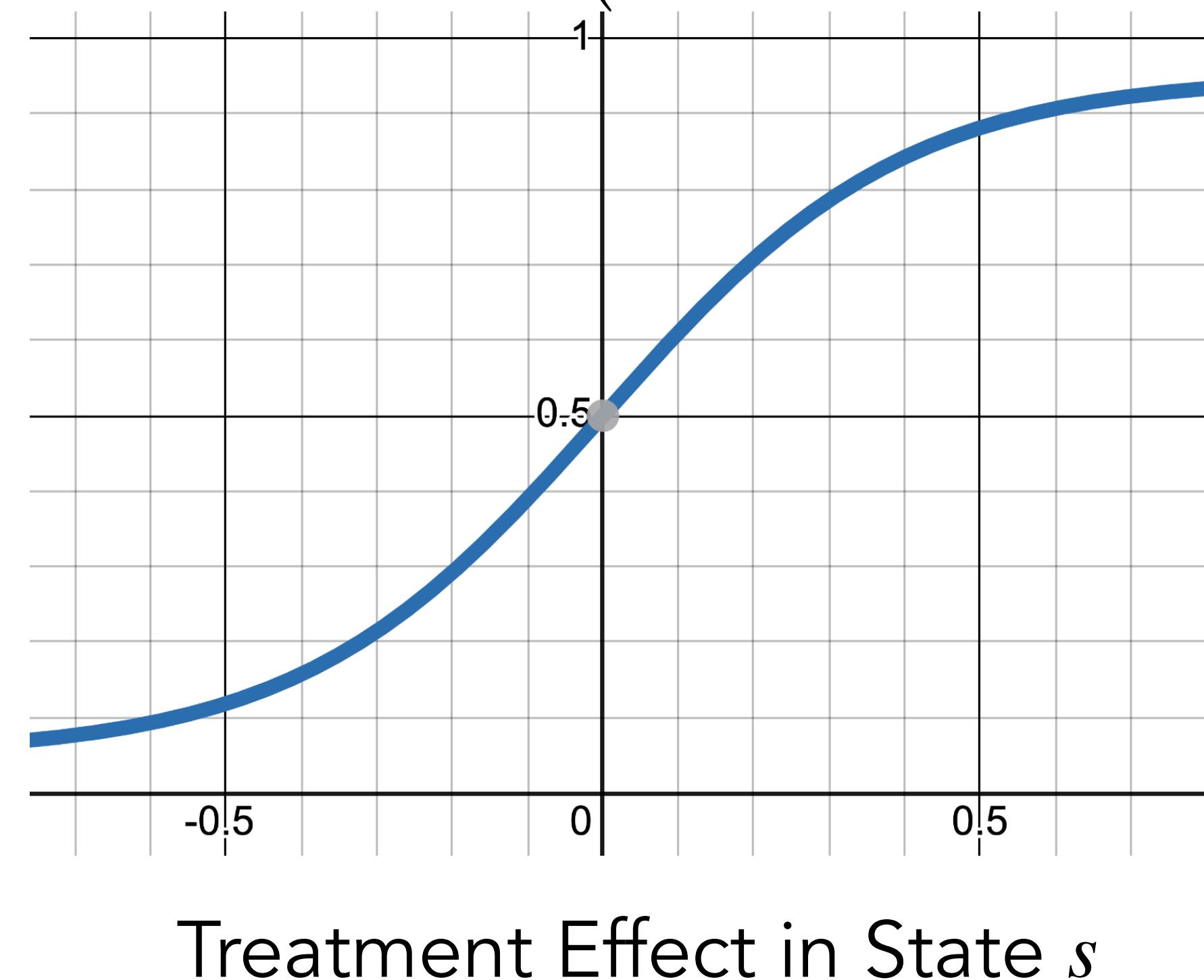
Treatment Effect in State  $s$

# Allocation Function: What probability should the limiting policy send a message?

**Balance Maximizing Rewards and Inferring Treatment Effects**  
- Between trial learning / Continual learning

$$\pi^*(s) = \text{Softmax}(\text{Treatment Effect}(s))$$

Probability of  
Sending a  
Message



No longer have issue of unstable learned policies from taking a "hardmax"

# Discussion Questions

(1) Explain the following:

- **(Our setting)** When a pooling RL algorithm that forms policies  $\{\hat{\pi}_t\}_{t=2}^T$  is used the resulting data trajectories  $H_{1,T}, H_{2,T}, \dots, H_{n,T}$  **are not independent** across people.
- **("Oracle" Setting)** When the target policies  $\{\pi_t^\star\}_{t=2}^T$  are used the resulting data trajectories  $H_{1,T}, H_{2,T}, \dots, H_{n,T}$  **are independent** across people.

(2) When do we expect using a pooling RL algorithm versus an individual RL algorithms to perform better? What are different ways to develop an RL algorithm that pools across users?

# Overview

1. Excursion effects after pooling
- 2. Overview of Inferential Approach**
3. Asymptotic Normality Proof Ideas

# Estimating Excursion Effects under Pooling

## Excursion Effect from Limiting Policy:

$$\mathbb{E}_{\pi^*} [Y_{t+1}(\bar{A}_{t-1}, 1) - Y_{t+1}(\bar{A}_{t-1}, 0) | X_t = x]$$

## Significant Challenges

- The data was collected under estimated policies  $\{\hat{\pi}_t\}_{t=1}^T$ , but we are interested in excursions from  $\{\pi_t^*\}_{t=1}^T$ 
  - We do not know the limiting policy  $\{\pi_t^*\}_{t=1}^T$ !!
- Our data trajectories are not independent across patients

# Estimating Excursion Effects under Pooling

Inferential target  $\theta^*$  solves:

$$0 = \mathbb{E}_{\pi^*} \left[ \frac{\partial}{\partial \theta} \ell(H_{i,T}; \theta^*) \right]$$

Set derivative of loss  
equal to zero to solve  
for minimizer

Estimator  $\hat{\theta}$  solves:

$$0 = \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \theta} \ell(H_{i,T}; \hat{\theta})$$

Example Least Squares Loss:  $\ell(H_{i,T}; \theta) = \sum_{t=1}^T (Y_{i,t+1} - X_{i,t}^\top \theta_0 - A_{i,t} \theta_1)^2$

# Estimating Excursion Effects under Pooling

Inferential target  $\theta^*$  solves:

$$0 = \mathbb{E}_{\pi^*} \left[ \frac{\partial}{\partial \theta} \ell(H_{i,T}; \theta^*) \right]$$

Set derivative of loss  
equal to zero to solve  
for minimizer

Estimator  $\hat{\theta}$  solves:

$$0 = \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \theta} \ell(H_{i,T}; \hat{\theta})$$

**Intuitively, why is forming estimators like this reasonable?**

- As  $n \rightarrow \infty$ , the policies  $\{\hat{\pi}_t\}_{t=1}^T$  will converge to the limiting policies  $\{\pi_t^*\}_{t=1}^T$
- Can include action centering, but omit for simplicity

What if you use standard inference approach on pooling online RL data?

**Can get confidence intervals that extremely overconfident!**

$$\ell(H_{i,T}; \theta) = \sum_{t=1}^T (Y_{i,t+1} - X_{i,t}^\top \theta_0 - A_{i,t} \theta_1)^2$$

**Coverage of 95% Confidence Intervals for Treatment Effect  $\theta_1^*$**

$\hat{\theta}_1$ Variance Estimators	$n = 50$	$n = 100$
Standard Sandwich	75.8%	77.6%

What if you use standard inference approach on pooling online RL data?

**Can get confidence intervals that extremely overconfident!**

$$\ell(H_{i,T}; \theta) = \sum_{t=1}^T (Y_{i,t+1} - X_{i,t}^\top \theta_0 - A_{i,t} \theta_1)^2$$

**Coverage of 95% Confidence Intervals for Treatment Effect  $\theta_1^*$**

$\hat{\theta}_1$ Variance Estimators	$n = 50$	$n = 100$
Standard Sandwich	75.8%	77.6%
“Adaptive” Sandwich	95.4%	96.5%

## Our “Adaptive” Sandwich Variance Estimator

- Data from pooling online RL algorithms → valid confidence intervals
- Applicable to inference for minimizers of general loss functions

Coverage of 95% Confidence Intervals for Treatment Effect  $\theta_1^*$

$\hat{\theta}_1$ Variance Estimators	$n = 50$	$n = 100$
Standard Sandwich	75.8%	77.6%
“Adaptive” Sandwich	95.4%	96.5%

# Impact of Adaptive Sandwich Variance Approach

Enables the use of pooling RL algorithms in digital intervention studies



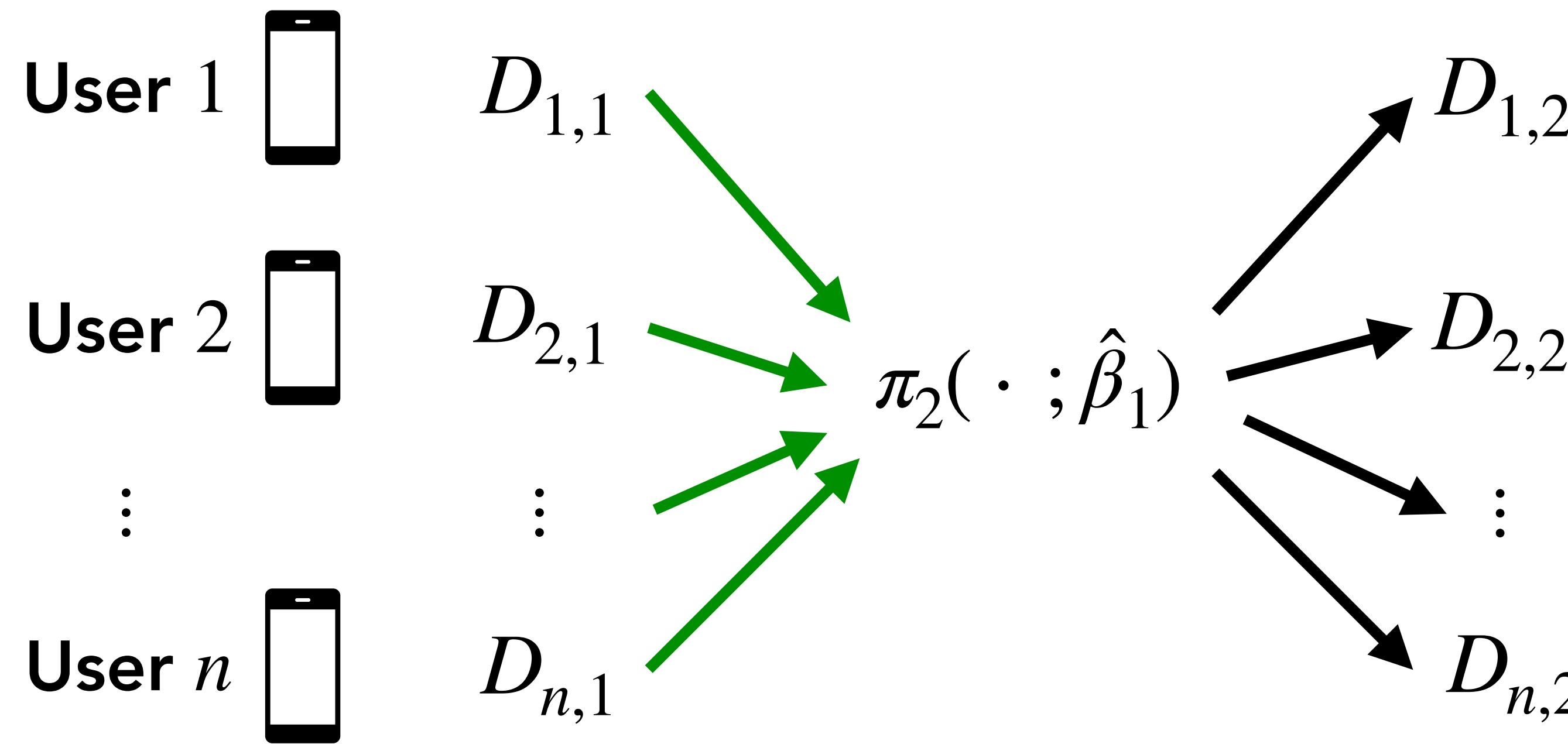
**Oralytics:**  
Oral Health Coaching



**MiWaves:**  
Curbing Adolescent Marijuana Use

# Inference Challenges

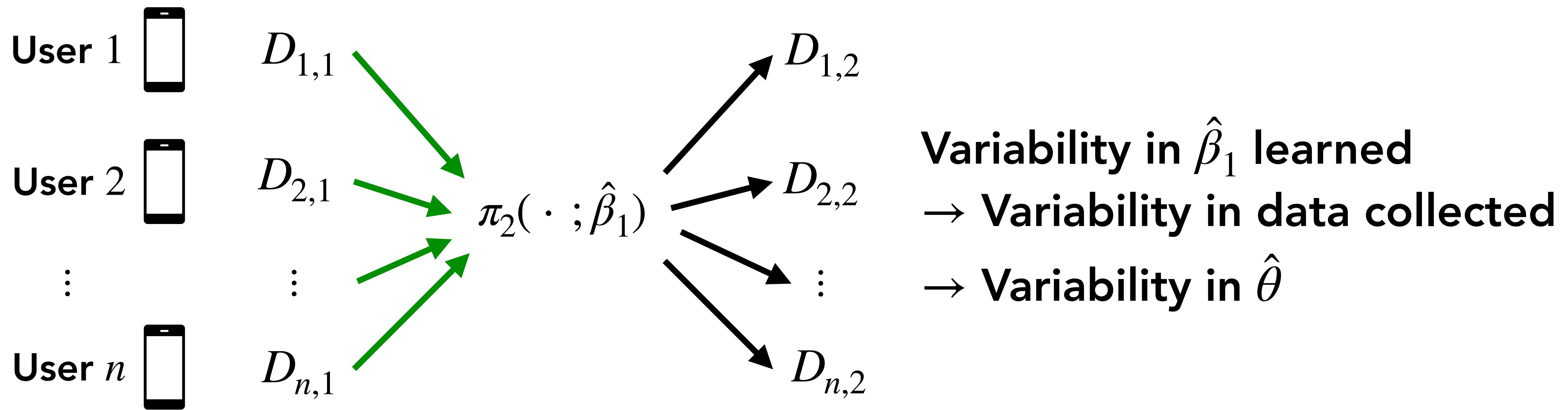
- (1) Dependencies both **within** and **between** users
- (2) Error of  $\hat{\theta}$  implicitly depends on how the algorithm forms and updates policies  $\hat{\pi}_t = \pi_t(\cdot ; \hat{\beta}_{t-1})$



**Variability in  $\hat{\beta}_1$  learned**  
→ **Variability in data collected**  
→ **Variability in  $\hat{\theta}$**

## Key Insight

Even though  $\{\hat{\beta}_t\}_{t=1}^{T-1}$  affect data collection if framed properly they can be mathematically treated like plug-in estimates of nuisance parameters that are used for data analysis.



# Standard Inference with “Plug-in” Nuisance Parameters

Given a dataset  $\{H_{i,T}\}_{i=1}^n$  where  $H_{1,T}, H_{2,T}, \dots, H_{n,T}$  are i.i.d.

(1) **Form a nuisance estimator  $\hat{\beta}$  that solves:**  $0 = \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \beta} g(H_{i,T}; \hat{\beta})$

(2) **“Plug-in”  $\beta = \hat{\beta}$  to solve for  $\hat{\theta}$  (data reuse):**  $0 = \frac{1}{N} \sum_{i=1}^N \frac{\partial}{\partial \theta} \ell(H_{i,T}; \hat{\theta}, \hat{\beta})$

Knowledge of how  $\hat{\theta}$  changes with different values of  $\hat{\beta}$  allows us to derive

joint asymptotic distribution:  $\sqrt{n} \begin{pmatrix} \hat{\beta} - \beta^\star \\ \hat{\theta} - \theta^\star \end{pmatrix} \xrightarrow{D} \mathcal{N}(0, \Sigma_{\theta,\beta})$

# Example: Observational Data Setting

Given a dataset  $\{H_{i,T}\}_{i=1}^n$  where  $H_{1,T}, H_{2,T}, \dots, H_{n,T}$  are i.i.d. collected by some unknown fixed policy

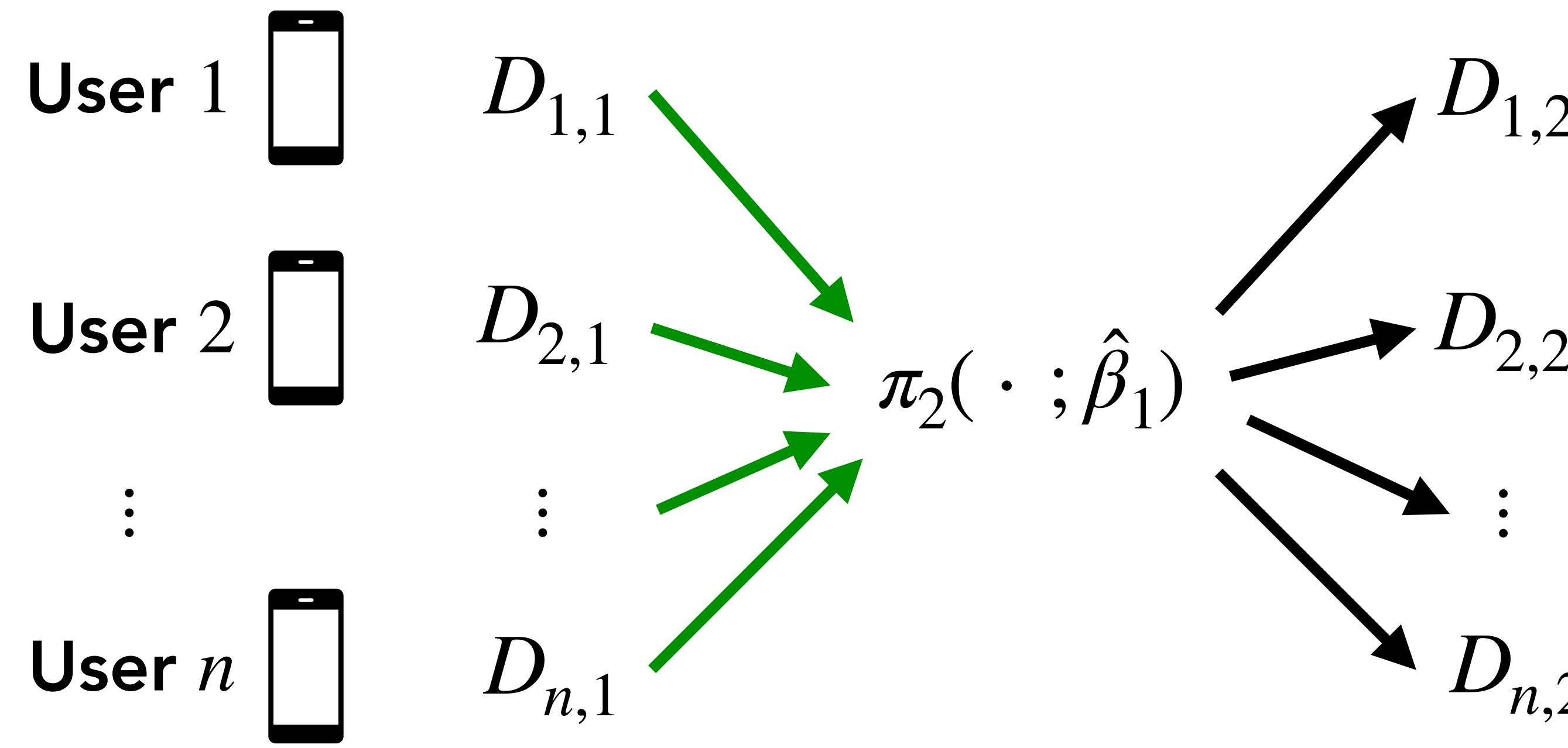
(1) **Form a nuisance estimator  $\hat{\beta}$  that solves:** 
$$0 = \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \beta} g(H_{i,T}; \hat{\beta})$$

$\hat{\beta}$  from fitted logistic regression model for  $\mathbb{P}(A_{i,t} = 1 | S_{i,t}) \approx \text{sigmoid}(S_{i,t}^\top \hat{\beta})$

(2) **“Plug-in”  $\beta = \hat{\beta}$  to solve for  $\hat{\theta}$  (data reuse):** 
$$0 = \frac{1}{N} \sum_{i=1}^N \frac{\partial}{\partial \theta} \ell(H_{i,T}; \hat{\theta}, \hat{\beta})$$

Forming estimator  $\hat{\theta}$  involves using the estimated action selection probabilities  $\text{sigmoid}(S_{i,t}^\top \hat{\beta})$

# In online RL setting...



$\hat{\beta}_1$  is not a plug-in estimator used to form  $\hat{\theta}$ .  
It is a property of the data collection procedure!!

$$\hat{\theta} \text{ solves } 0 = \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \theta} \ell(H_{i,T}; \hat{\theta})$$

# Importance Weights as a Theoretical Tool

**"Simple" Solution:**  $\hat{\theta}$  solves for  $\{\beta_t\}_{t=1}^{T-1} = \{\hat{\beta}_t\}_{t=1}^{T-1}$

$$0 = \frac{1}{n} \sum_{i=1}^n \left\{ \prod_{t=2}^T \left( \frac{\pi_t(S_{i,t}; \beta_{t-1})}{\pi_t(S_{i,t}; \hat{\beta}_{t-1})} \right)^{A_{i,t}} \left( \frac{1 - \pi_t(S_{i,t}; \beta_{t-1})}{1 - \pi_t(S_{i,t}; \hat{\beta}_{t-1})} \right)^{1-A_{i,t}} \right\} \frac{\partial}{\partial \theta} \ell(H_{i,T}; \hat{\theta})$$

- Weights are not used to form  $\hat{\theta}!!$
- Allows us to capture how changes in  $\{\beta_t\}_{t=2}^T$  affect errors in  $\hat{\theta}$   
→ analyze similarly to plug-in estimators of nuisance parameters

# Adaptive Sandwich Variance

$$\sqrt{n}(\hat{\theta}^{(n)} - \theta^\star) \xrightarrow{D} \mathcal{N}(0, \ddot{L}_\theta^{-1} \Sigma^{\text{adapt}} \ddot{L}_\theta^{-1})$$

$$\Sigma^{\text{adapt}} = \mathbb{E}_{\pi^\star} \left[ \left\{ \frac{\partial}{\partial \theta} \ell(H_{i,T}; \theta^\star) + \underbrace{\sum_{t=1}^{T-1} M_t \dot{g}_t(H_{i,t}; \beta_t^\star)}_{\text{Correction in Variance Due to Pooled RL Algorithm}} \right\}^{\otimes 2} \right]$$

$$\ddot{L}_\theta = \mathbb{E}_{\pi^\star} \left[ \frac{\partial^2}{\partial \theta \partial \theta} \ell(H_{i,T}; \theta) \right]$$

$M_t$  given in paper: Statistical Inference After Adaptive Sampling for Longitudinal Data  
(<https://arxiv.org/abs/2202.07098>)

# Asymptotic Normality Proof Ideas

# Overview

1. Proving normality in “oracle” setting

( $H_{1,T}, H_{2,T}, \dots, H_{n,T}$  are i.i.d.)

2. Proving normality when  $H_{1,T}, H_{2,T}, \dots, H_{n,T}$  are  
collected with a pooling RL algorithm ( $T = 2$  case)

# “Oracle” Setting with i.i.d. Data Trajectories

- Given a dataset  $\{H_{i,T}\}_{i=1}^n$  where  $H_{1,T}, H_{2,T}, \dots, H_{n,T}$  are i.i.d. collected by known target policies  $\{\pi_t^\star\}_{t=1}^T$
- Estimand  $\theta^\star$  where  $\theta = \theta^\star$  solves

$$0 = \mathbb{E}_{\pi^\star} [\dot{\ell}_i(\theta)] \triangleq \mathbb{E}_{\pi^\star} \left[ \frac{\partial}{\partial \theta} \ell(H_{i,T}; \theta) \right]$$

- Estimator  $\hat{\theta}$  where  $\theta = \hat{\theta}$  solves

$$0 = \mathbb{P}_n [\dot{\ell}_i(\theta)] \triangleq \mathbb{E}_{\pi^\star} \left[ \frac{\partial}{\partial \theta} \ell(H_{i,T}; \theta) \right]$$

Example Least Squares Loss:  $\ell(H_{i,T}; \theta) = \sum_{t=1}^T (Y_{i,t+1} - X_{i,t}^\top \theta_0 - A_{i,t} \theta_1)^2$

# Normality Result (Standard Sandwich Variance)

$$\sqrt{n} \left( \hat{\theta} - \theta^* \right) \xrightarrow{D} \mathcal{N} \left( 0, \ddot{L}^{-1} \Sigma (\dot{L}^{-1})^\top \right)$$

where

$$\Sigma = \mathbb{E}_{\pi^*} \left[ \dot{\ell}_i(\theta^*) \dot{\ell}_i(\theta^*)^\top \right]$$

Following Theorem  
5.21 of Van Der Vaart,  
Asymptotic Statistics

and

$$\ddot{L} = \frac{\partial}{\partial \theta} \mathbb{E}_{\pi^*} \left[ \dot{\ell}_i(\theta) \right] \Big|_{\theta=\theta^*} = \mathbb{E}_{\pi^*} \left[ \frac{\partial^2}{\partial \theta \partial \theta} \ell_i(\theta) \right] \Big|_{\theta=\theta^*}$$

# Proof Outline

$$\Sigma = \mathbb{E}_{\pi^\star} \left[ \dot{\ell}_i(\theta^\star) \dot{\ell}_i(\theta^\star)^\top \right]$$

$$\sqrt{n}(\hat{\theta} - \theta^\star) \xrightarrow{D} \mathcal{N}(0, \ddot{L}^{-1}\Sigma(\ddot{L}^{-1})^\top)$$

$$(1) \sqrt{n}(\mathbb{P}_n \dot{\ell}_i(\theta^\star) - \mathbb{E}_{\pi^\star}[\dot{\ell}_i(\theta^\star)]) \xrightarrow{D} \mathcal{N}(0, \Sigma)$$

$$(2) \sqrt{n}(\mathbb{P}_n \dot{\ell}_i(\theta^\star) - \mathbb{E}_{\pi^\star}[\dot{\ell}_i(\theta^\star)])$$

$$= -\ddot{L}\sqrt{n}(\hat{\theta} - \theta^\star) + \sqrt{n}o_P(\|\hat{\theta} - \theta^\star\|_2) + o_P(1)$$

$$(3) \sqrt{n}o_P(\|\hat{\theta} - \theta^\star\|_2) = o_P(1)$$

# Proof Outline

$$\Sigma = \mathbb{E}_{\pi^\star} \left[ \dot{\ell}_i(\theta^\star) \dot{\ell}_i(\theta^\star)^\top \right]$$

$$\sqrt{n}(\hat{\theta} - \theta^\star) \xrightarrow{D} \mathcal{N}(0, \ddot{L}^{-1}\Sigma(\ddot{L}^{-1})^\top)$$

$$(1) \sqrt{n}(\mathbb{P}_n \dot{\ell}_i(\theta^\star) - \mathbb{E}_{\pi^\star}[\dot{\ell}_i(\theta^\star)]) \xrightarrow{D} \mathcal{N}(0, \Sigma)$$

Central Limit  
Theorem

$$(2) \sqrt{n}(\mathbb{P}_n \dot{\ell}_i(\theta^\star) - \mathbb{E}_{\pi^\star}[\dot{\ell}_i(\theta^\star)])$$

$$= -\ddot{L}\sqrt{n}(\hat{\theta} - \theta^\star) + \sqrt{n}o_P(\|\hat{\theta} - \theta^\star\|_2) + o_P(1)$$

$$(3) \sqrt{n}o_P(\|\hat{\theta} - \theta^\star\|_2) = o_P(1)$$

# Proof Outline

$$\Sigma = \mathbb{E}_{\pi^\star} \left[ \dot{\ell}_i(\theta^\star) \dot{\ell}_i(\theta^\star)^\top \right]$$

$$\sqrt{n}(\hat{\theta} - \theta^\star) \xrightarrow{D} \mathcal{N}(0, \ddot{L}^{-1}\Sigma(\ddot{L}^{-1})^\top)$$

$$(1) \sqrt{n}(\mathbb{P}_n \dot{\ell}_i(\theta^\star) - \mathbb{E}_{\pi^\star}[\dot{\ell}_i(\theta^\star)]) \xrightarrow{D} \mathcal{N}(0, \Sigma)$$

$$(2) \sqrt{n}(\mathbb{P}_n \dot{\ell}_i(\theta^\star) - \mathbb{E}_{\pi^\star}[\dot{\ell}_i(\theta^\star)]) \\ = -\ddot{L}\sqrt{n}(\hat{\theta} - \theta^\star) + \sqrt{n}o_P(\|\hat{\theta} - \theta^\star\|_2) + o_P(1)$$

$$(3) \sqrt{n}o_P(\|\hat{\theta} - \theta^\star\|_2) = o_P(1)$$

# Notation Slide

- A sequence of random variables  $Z_n = o_P(1)$  if for any  $\epsilon > 0$ , as

$$n \rightarrow \infty, P(\|Z_n\|_2 > \epsilon) \rightarrow 0$$

- More generally,  $Z_n = o_P(B_n)$  for some random sequence  $B_n$ , if for

$$\text{any } \epsilon > 0, \text{ as } n \rightarrow \infty, P\left(\frac{\|Z_n\|_2}{\|B_n\|_2} > \epsilon\right) \rightarrow 0$$

- See Van der Vaart, Asymptotic Statistics, Chapter 2.2

## Step (2) Outline

$$\sqrt{n} \left( \mathbb{P}_n \dot{\ell}_i(\theta^*) - \mathbb{E}_{\pi^*} [\dot{\ell}_i(\theta^*)] \right)$$

$$= - \ddot{L} \sqrt{n} (\hat{\theta} - \theta^*) + \sqrt{n} o_P(\|\hat{\theta} - \theta^*\|_2) + o_P(1)$$

## Step (2) Outline

## Asymptotic Equicontinuity

$$\sqrt{n}(\mathbb{P}_n \dot{\ell}_i(\theta^*) - \mathbb{E}_{\pi^*}[\dot{\ell}_i(\theta^*)])$$

$$= \sqrt{n}(\mathbb{P}_n \dot{\ell}_i(\hat{\theta}) - \mathbb{E}_{\pi^*}[\dot{\ell}_i(\hat{\theta})]) + o_P(1)$$

$$= \sqrt{n}(\mathbb{E}_{\pi^*}[\dot{\ell}_i(\theta^*)] - \mathbb{E}_{\pi^*}[\dot{\ell}_i(\hat{\theta})]) + o_P(1)$$

$$= -\ddot{L}\sqrt{n}(\hat{\theta} - \theta^*) + \sqrt{n}o_P(\|\hat{\theta} - \theta^*\|_2) + o_P(1)$$

- Use  $\hat{\theta} \xrightarrow{P} \theta^*$
- Show that random mapping is continuous in  $\theta$
- Apply continuous mapping theorem

Why does the following equality hold?

$$\sqrt{n}(\mathbb{P}_n \dot{\ell}_i(\theta^*) - \mathbb{E}_{\pi^*}[\dot{\ell}_i(\theta^*)])$$

Using Definitions  
of  $\hat{\theta}$ ,  $\theta^*$

$$= \sqrt{n}(\mathbb{P}_n \dot{\ell}_i(\hat{\theta}) - \mathbb{E}_{\pi^*}[\dot{\ell}_i(\hat{\theta})]) + o_P(1)$$

$\hat{\theta}$  solves  $0 = \mathbb{P}_n \dot{\ell}_i(\hat{\theta})$

$$= \sqrt{n}(\mathbb{E}_{\pi^*}[\dot{\ell}_i(\theta^*)] - \mathbb{E}_{\pi^*}[\dot{\ell}_i(\hat{\theta})]) + o_P(1)$$

$\theta^*$  solves

$$0 = \mathbb{E}_{\pi^*}[\dot{\ell}_i(\theta^*)]$$

$$= -\ddot{L}\sqrt{n}(\hat{\theta} - \theta^*) + \sqrt{n}o_P(\|\hat{\theta} - \theta^*\|_2) + o_P(1)$$

Why does the following equality hold?

$$\sqrt{n}(\mathbb{P}_n \dot{\ell}_i(\theta^*) - \mathbb{E}_{\pi^*}[\dot{\ell}_i(\theta^*)])$$

Differentiability

$$= \sqrt{n}(\mathbb{P}_n \dot{\ell}_i(\hat{\theta}) - \mathbb{E}_{\pi^*}[\dot{\ell}_i(\hat{\theta})]) + o_P(1)$$

$$\ddot{L} = \frac{\partial}{\partial \theta} \mathbb{E}_{\pi^*} [\dot{\ell}_i(\theta)] \Big|_{\theta=\theta^*}$$

$$= \sqrt{n}(\mathbb{E}_{\pi^*}[\dot{\ell}_i(\theta^*)] - \mathbb{E}_{\pi^*}[\dot{\ell}_i(\hat{\theta})]) + o_P(1)$$

$$= -\ddot{L}\sqrt{n}(\hat{\theta} - \theta^*) + \sqrt{n}o_P(\|\hat{\theta} - \theta^*\|_2) + o_P(1)$$

# Differentiability

$$\ddot{L} = \frac{\partial}{\partial \theta} \mathbb{E}_{\pi^\star} \left[ \dot{\ell}_i(\theta) \right] \Big|_{\theta=\theta^\star}$$

$$\lim_{\theta \rightarrow \theta^\star} \frac{\left\| \mathbb{E}_{\pi^\star} \left[ \dot{\ell}_i(\theta) \right] - \mathbb{E}_{\pi^\star} \left[ \dot{\ell}_i(\theta^\star) \right] - \ddot{L}(\theta - \theta^\star) \right\|_2}{\|\theta - \theta^\star\|_2} = 0$$

$$\mathbb{E}_{\pi^\star} \left[ \dot{\ell}_i(\hat{\theta}) \right] - \mathbb{E}_{\pi^\star} \left[ \dot{\ell}_i(\theta^\star) \right] - \ddot{L}(\hat{\theta} - \theta^\star) = o_P \left( \|\hat{\theta} - \theta^\star\|_2 \right)$$

$$\begin{aligned} \sqrt{n} \left( \mathbb{E}_{\pi^\star} \left[ \dot{\ell}_i(\theta^\star) \right] - \mathbb{E}_{\pi^\star} \left[ \dot{\ell}_i(\hat{\theta}) \right] \right) + o_P(1) \\ = - \ddot{L} \sqrt{n} (\hat{\theta} - \theta^\star) + \sqrt{n} o_P \left( \|\hat{\theta} - \theta^\star\|_2 \right) + o_P(1) \end{aligned}$$

# Proof Outline

$$\Sigma = \mathbb{E}_{\pi^\star} \left[ \dot{\ell}_i(\theta^\star) \dot{\ell}_i(\theta^\star)^\top \right]$$

$$\sqrt{n}(\hat{\theta} - \theta^\star) \xrightarrow{D} \mathcal{N}(0, \ddot{L}^{-1}\Sigma(\ddot{L}^{-1})^\top)$$

$$(1) \sqrt{n}(\mathbb{P}_n \dot{\ell}_i(\theta^\star) - \mathbb{E}_{\pi^\star}[\dot{\ell}_i(\theta^\star)]) \xrightarrow{D} \mathcal{N}(0, \Sigma)$$

$$(2) \sqrt{n}(\mathbb{P}_n \dot{\ell}_i(\theta^\star) - \mathbb{E}_{\pi^\star}[\dot{\ell}_i(\theta^\star)])$$

$$= -\ddot{L}\sqrt{n}(\hat{\theta} - \theta^\star) + \sqrt{n}o_P(\|\hat{\theta} - \theta^\star\|_2) + o_P(1)$$

$$(3) \sqrt{n}o_P(\|\hat{\theta} - \theta^\star\|_2) = o_P(1)$$

HW Exercise (uses above two steps)

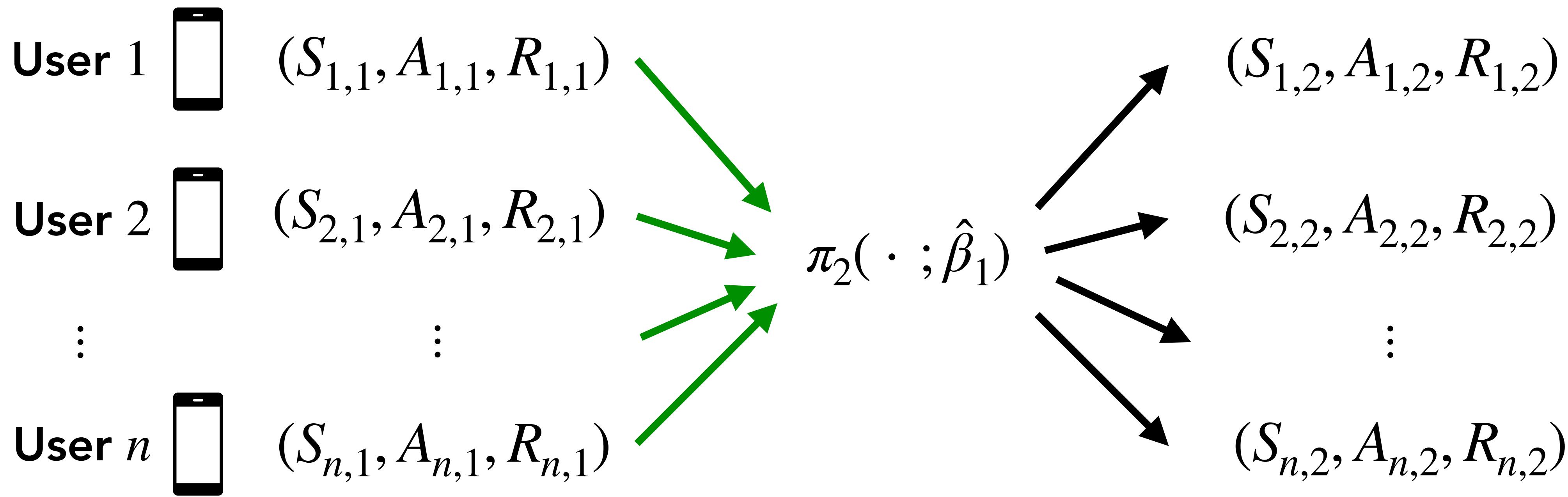
# Overview

1. Proving normality in “oracle” setting

( $H_{1,T}, H_{2,T}, \dots, H_{n,T}$  are i.i.d.)

2. Proving normality when  $H_{1,T}, H_{2,T}, \dots, H_{n,T}$  are  
collected with a pooling RL algorithm ( $T = 2$  case)

# Recap of Problem Setting $T = 2$



- Use  $\pi_1$  to collect  $\{(S_{i,1}, A_{i,1}, R_{i,1})\}_{i=1}^n$
- Use  $\{(S_{i,1}, A_{i,1}, R_{i,1})\}_{i=1}^n$  to form  $\hat{\beta}_1^{(n)}$
- Use  $\hat{\pi}_2(\cdot) = \pi(\cdot ; \hat{\beta}_1^{(n)})$  to collect  $\{(S_{i,2}, A_{i,2}, R_{i,2})\}_{i=1}^n$

# Pooling Setting with non-i.i.d. Data Trajectories

- Given a dataset  $\{H_{i,2}\}_{i=1}^n$  collected by estimated policy  $\hat{\pi}_2$
- Estimand  $\theta^\star$  where  $\theta = \theta^\star$  solves

$$0 = \mathbb{E}_{\pi^\star} [\dot{\ell}_i(\theta)] \triangleq \mathbb{E}_{\pi^\star} \left[ \frac{\partial}{\partial \theta} \ell(H_{i,2}; \theta) \right]$$

- Estimator  $\hat{\theta}$  where  $\theta = \hat{\theta}$  solves

$$0 = \mathbb{P}_n [\dot{\ell}_i(\theta)] \triangleq \mathbb{E}_{\pi^\star} \left[ \frac{\partial}{\partial \theta} \ell(H_{i,2}; \theta) \right]$$

Example Least Squares Loss:  $\ell(H_{i,T}; \theta) = \sum_{t=1}^T (Y_{i,t+1} - X_{i,t}^\top \theta_0 - A_{i,t} \theta_1)^2$

# $\hat{\beta}_1$ as Estimator of Nuisance

Inferential target  $\theta^*$  solves:

$$0 = \mathbb{E}_{\pi^*} \left[ \dot{\ell}_i(\theta^*) \right]$$

- $\pi_2^* = \pi_2(\cdot; \beta_1^*)$  is a nuisance function estimated with  
 $\hat{\pi}_2^{(n)} = \pi_2(\cdot; \hat{\beta}_1^{(n)})$
- **However, nuisance function is estimated by the RL algorithm rather than the data analyst**

# $T = 2$ Case: Loss Function for $\beta_1$

- Formed by the RL algorithm
- Limiting  $\beta_1^*$

$$0 = \mathbb{E}_{\pi^*}[\dot{g}_{i,1}(\beta_1^*)] = \mathbb{E}_{\pi^*}\left[\frac{\partial}{\partial \beta} g_1(H_{i,1}; \beta) \Big|_{\beta=\beta_1^*}\right]$$

- Estimator  $\hat{\beta}_1$

$$0 = \mathbb{P}_n \dot{g}_{i,1}(\hat{\beta}_1) = \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \beta} g_1(H_{i,1}; \beta) \Big|_{\beta=\beta_1^*}$$

# Normality Result (Adaptive Sandwich Variance)

$$\sqrt{n} \left( \hat{\theta} - \theta^* \right) \xrightarrow{D} \mathcal{N} \left( 0, \ddot{L}^{-1} \Sigma^{\text{adapt}} (\ddot{L}^{-1})^\top \right)$$

where

$$\Sigma^{\text{adapt}} = \mathbb{E}_{\pi^*} \left[ \left\{ \dot{\ell}_i(\theta^*) - V_1 \ddot{G}_1^{-1} \dot{g}_{i,1}(\beta_1^*) \right\} \left\{ \dot{\ell}_i(\theta^*) - V_1 \ddot{G}_1^{-1} \dot{g}_{i,1}(\beta_1^*) \right\}^\top \right]$$

$$\ddot{G}_1 = \frac{\partial}{\partial \beta_1} \mathbb{E} \left[ \dot{g}_{1,i}(\beta_1) \right] \Big|_{\beta_1=\beta_1^*}$$

and

$$\ddot{L} = \frac{\partial}{\partial \theta} \mathbb{E}_{\pi^*} \left[ \dot{\ell}_i(\theta) \right] \Big|_{\theta=\theta^*}$$

# Joint Asymptotic Normality Result

$$\sqrt{n} \begin{pmatrix} \hat{\beta}_1 - \beta_1^* \\ \hat{\theta} - \theta^* \end{pmatrix} \xrightarrow{D} \mathcal{N} \left( 0, \begin{bmatrix} \ddot{G}_1 & 0 \\ V_1 & \ddot{L} \end{bmatrix}^{-1} \Sigma_{1:2} \begin{bmatrix} \ddot{G}_1 & 0 \\ V_1 & \ddot{L} \end{bmatrix}^{-1, \top} \right)$$

$$\Sigma_{1:2} \triangleq \mathbb{E}_{\pi^*} \left[ \begin{pmatrix} \dot{g}_i(\beta_1^*) \\ \dot{\ell}_i(\theta^*) \end{pmatrix} \begin{pmatrix} \dot{g}_i(\beta_1^*) \\ \dot{\ell}_i(\theta^*) \end{pmatrix}^\top \right] \text{ and } V_1 = \frac{\partial}{\partial \beta_1} \mathbb{E}_{\pi_2(\beta_1)} [\dot{\ell}_i(\theta^*)] \Big|_{\beta_1=\beta_1^*}$$

# Joint Asymptotic Normality Result

$$\sqrt{n} \begin{pmatrix} \hat{\beta}_1 - \beta_1^* \\ \hat{\theta} - \theta^* \end{pmatrix} \xrightarrow{D} \mathcal{N} \left( 0, \begin{bmatrix} \ddot{G}_1 & 0 \\ V_1 & \ddot{L} \end{bmatrix}^{-1} \Sigma_{1:2} \begin{bmatrix} \ddot{G}_1 & 0 \\ V_1 & \ddot{L} \end{bmatrix}^{-1, \top} \right)$$

$$\Sigma_{1:2} \triangleq \mathbb{E}_{\pi^*} \left[ \begin{pmatrix} \dot{g}_i(\beta_1^*) \\ \dot{\ell}_i(\theta^*) \end{pmatrix} \begin{pmatrix} \dot{g}_i(\beta_1^*) \\ \dot{\ell}_i(\theta^*) \end{pmatrix}^\top \right] \text{ and } V_1 = \frac{\partial}{\partial \beta_1} \mathbb{E}_{\pi_2(\beta_1)} [\dot{\ell}_i(\theta^*)] \Big|_{\beta_1=\beta_1^*}$$

**Interpretation of  $V_1$ :** Change in criterion for  $\theta^*$  with little changes in policy used to collect data (i.e.  $\beta_1$ )

# $T = 2$ Setting: Naive Approach

**Joint Criteria for  $\hat{\beta}_1, \hat{\theta}$ :**

$$0 = \frac{1}{n} \sum_{i=1}^n \begin{pmatrix} \dot{g}_{i,1}(\hat{\beta}_1) \\ \dot{\ell}_i(\hat{\theta}) \end{pmatrix}$$

- **Issue:** Above the relationship between  $\hat{\theta}$  and  $\hat{\beta}_1$  is not explicit
  - We use weighting to represent how  $\hat{\theta}$  is affected by estimation of the nuisance  $\pi_2^\star = \pi_2(\cdot; \beta_1^\star)$

# $T = 2$ Setting: Radon-Nikodym Weights

**Joint Criteria for  $\hat{\beta}_1, \hat{\theta}$ :**

$$0 = \frac{1}{n} \sum_{i=1}^n \begin{pmatrix} \dot{g}_{i,1}(\hat{\beta}_1) \\ W_{i,2}(\hat{\beta}_1) \dot{\ell}_i(\hat{\theta}) \end{pmatrix}$$

$\hat{\beta}_1$  is also a plug-in into the Radon-Nikodym weight!

where

$$W_{i,2}(\beta_1) \triangleq \left( \frac{\pi_{i,2}(\beta_1)}{\pi_{i,2}(\hat{\beta}_1)} \right)^{A_{i,2}} \left( \frac{1 - \pi_{i,2}(\beta_1)}{1 - \pi_{i,2}(\hat{\beta}_1)} \right)^{1-A_{i,2}}$$

$$\pi_{i,2}(\beta_1) \triangleq \pi_2(S_{i,2}; \beta_1)$$

# Proof Key Steps

$$\sqrt{n} \begin{pmatrix} \hat{\beta}_1 - \beta_1^* \\ \hat{\theta} - \theta^* \end{pmatrix} \xrightarrow{D} \mathcal{N}\left(0, \begin{bmatrix} \ddot{G}_1 & 0 \\ V_1 & \ddot{L} \end{bmatrix}^{-1} \Sigma_{1:2} \begin{bmatrix} \ddot{G}_1 & 0 \\ V_1 & \ddot{L} \end{bmatrix}^{-1, \top}\right)$$

## Proof Key Steps

$$\sqrt{n} \begin{pmatrix} \hat{\beta}_1 - \beta_1^* \\ \hat{\theta} - \theta^* \end{pmatrix} \xrightarrow{D} \mathcal{N}\left(0, \begin{bmatrix} \ddot{G}_1 & 0 \\ V_1 & \ddot{L} \end{bmatrix}^{-1} \Sigma_{1:2} \begin{bmatrix} \ddot{G}_1 & 0 \\ V_1 & \ddot{L} \end{bmatrix}^{-1, \top} \right)$$

$$(1) \sqrt{n} \left( \mathbb{P}_n \begin{bmatrix} \dot{g}_i(\beta_1^*) \\ W_{i,1}(\beta_1^*) \dot{\ell}_i(\theta^*) \end{bmatrix} - \mathbb{E} \begin{bmatrix} \dot{g}_i(\beta_1^*) \\ W_{i,1}(\beta_1^*) \dot{\ell}_i(\theta^*) \end{bmatrix} \right) \xrightarrow{D} \mathcal{N}(0, \Sigma_{1:2})$$

Our data is no longer independent across users!

Weighted Martingale CLT

# Proof Key Steps

$$\sqrt{n} \begin{pmatrix} \hat{\beta}_1 - \beta_1^* \\ \hat{\theta} - \theta^* \end{pmatrix} \xrightarrow{D} \mathcal{N} \left( 0, \begin{bmatrix} \ddot{G}_1 & 0 \\ V_1 & \ddot{L} \end{bmatrix}^{-1} \Sigma_{1:2} \begin{bmatrix} \ddot{G}_1 & 0 \\ V_1 & \ddot{L} \end{bmatrix}^{-1, \top} \right)$$

$$\begin{aligned}
& (2) \sqrt{n} \left( \mathbb{P}_n \begin{bmatrix} \dot{g}_i(\beta_1^*) \\ W_{i,1}(\beta_1^*) \dot{\ell}_i(\theta^*) \end{bmatrix} - \mathbb{E} \begin{bmatrix} \dot{g}_i(\beta_1^*) \\ W_{i,1}(\beta_1^*) \dot{\ell}_i(\theta^*) \end{bmatrix} \right) \\
& = \begin{bmatrix} \ddot{G}_1 & 0 \\ V_1 & \ddot{L} \end{bmatrix}^{-1} \sqrt{n} \begin{pmatrix} \hat{\beta}_1 - \beta_1^* \\ \hat{\theta} - \theta^* \end{pmatrix} + \sqrt{n} o_P \left( \left\| \begin{pmatrix} \hat{\beta}_1 - \beta_1^* \\ \hat{\theta} - \theta^* \end{pmatrix} \right\| \right) + o_P(1)
\end{aligned}$$

## Proof Outline of Step (2)

$$\sqrt{n} \left( \mathbb{P}_n \begin{bmatrix} \dot{g}_i(\beta_1^*) \\ W_{i,1}(\beta_1^*) \dot{\ell}_i(\theta^*) \end{bmatrix} - \mathbb{E} \begin{bmatrix} \dot{g}_i(\beta_1^*) \\ W_{i,1}(\beta_1^*) \dot{\ell}_i(\theta^*) \end{bmatrix} \right)$$

$$= \begin{bmatrix} \ddot{G}_1 & 0 \\ V_1 & \ddot{L} \end{bmatrix}^{-1} \sqrt{n} \begin{pmatrix} \hat{\beta}_1 - \beta_1^* \\ \hat{\theta} - \theta^* \end{pmatrix} + \sqrt{n} o_P \left( \begin{array}{c} \|\hat{\beta}_1 - \beta_1^*\| \\ \|\hat{\theta} - \theta^*\| \end{array} \right) + o_P(1)$$

## Proof Outline of Step (2)

$$\sqrt{n} \left( \mathbb{P}_n \begin{bmatrix} \dot{g}_i(\beta_1^*) \\ W_{i,1}(\beta_1^*) \dot{\ell}_i(\theta^*) \end{bmatrix} - \mathbb{E} \begin{bmatrix} \dot{g}_i(\beta_1^*) \\ W_{i,1}(\beta_1^*) \dot{\ell}_i(\theta^*) \end{bmatrix} \right)$$

$$= \sqrt{n} \left( \mathbb{P}_n \begin{bmatrix} \dot{g}_i(\hat{\beta}_1) \\ W_{i,1}(\hat{\beta}_1) \dot{\ell}_i(\hat{\theta}) \end{bmatrix} - \mathbb{E} \begin{bmatrix} \dot{g}_i(\hat{\beta}_1) \\ W_{i,1}(\hat{\beta}_1) \dot{\ell}_i(\hat{\theta}) \end{bmatrix} \right) + o_P(1)$$

$$= \begin{bmatrix} \ddot{G}_1 & 0 \\ V_1 & \ddot{L} \end{bmatrix}^{-1} \sqrt{n} \begin{pmatrix} \hat{\beta}_1 - \beta_1^* \\ \hat{\theta} - \theta^* \end{pmatrix} + \left( \begin{array}{c} \|\hat{\beta}_1 - \beta_1^*\| \\ \|\hat{\theta} - \theta^*\| \end{array} \right) + o_P(1)$$

Asymptotic  
Equicontinuity

- Use that  $(\hat{\beta}_1, \hat{\theta}) \xrightarrow{P} (\beta_1^*, \theta^*)$
- Show that random mapping is continuous in  $(\beta_1, \theta)$
- Apply continuous mapping theorem

# Why does this equality hold?

$$\sqrt{n} \left( \mathbb{P}_n \begin{bmatrix} \dot{g}_i(\beta^\star) \\ W_{i,1}(\beta_1^\star) \dot{\ell}_i(\theta^\star) \end{bmatrix} - \mathbb{E} \begin{bmatrix} \dot{g}_i(\beta_1^\star) \\ W_{i,1}(\beta_1^\star) \dot{\ell}_i(\theta^\star) \end{bmatrix} \right)$$

Using Definitions of  
 $\hat{\beta}_1, \hat{\theta}, \beta_1^\star, \theta^\star$

$$\begin{aligned} &= \sqrt{n} \left( \mathbb{P}_n \begin{bmatrix} \dot{g}_i(\hat{\beta}_1) \\ W_{i,1}(\hat{\beta}_1) \dot{\ell}_i(\hat{\theta}) \end{bmatrix} - \mathbb{E} \begin{bmatrix} \dot{g}_i(\hat{\beta}_1) \\ W_{i,1}(\hat{\beta}_1) \dot{\ell}_i(\hat{\theta}) \end{bmatrix} \right) + o_P(1) \\ &= \sqrt{n} \left( \mathbb{E} \begin{bmatrix} \dot{g}_i(\beta_1^\star) \\ W_{i,1}(\beta_1^\star) \dot{\ell}_i(\theta^\star) \end{bmatrix} - \mathbb{E} \begin{bmatrix} \dot{g}_i(\hat{\beta}_1) \\ W_{i,1}(\hat{\beta}_1) \dot{\ell}_i(\hat{\theta}) \end{bmatrix} \right) + o_P(1) \end{aligned}$$

$$= \begin{bmatrix} \ddot{G}_1 & 0 \\ V_1 & \ddot{L} \end{bmatrix}^{-1} \sqrt{n} \begin{pmatrix} \hat{\beta}_1 - \beta_1^\star \\ \hat{\theta} - \theta^\star \end{pmatrix} + \left( \begin{array}{c} \|\hat{\beta}_1 - \beta_1^\star\| \\ \|\hat{\theta} - \theta^\star\| \end{array} \right) + o_P(1)$$

$\hat{\beta}_1, \hat{\theta}$  solves

$$0 = \mathbb{P}_n \begin{bmatrix} \dot{g}_i(\hat{\beta}_1) \\ W_{i,1}(\hat{\beta}_1) \dot{\ell}_i(\hat{\theta}) \end{bmatrix}$$

$\theta^\star$  solves

$$0 = \mathbb{E} \begin{bmatrix} \dot{g}_i(\beta_1^\star) \\ W_{i,1}(\beta_1^\star) \dot{\ell}_i(\theta^\star) \end{bmatrix}$$

# Why does this equality hold?

$$\sqrt{n} \left( \mathbb{P}_n \begin{bmatrix} \dot{g}_i(\beta^\star) \\ W_{i,1}(\beta_1^\star) \dot{\ell}_i(\theta^\star) \end{bmatrix} - \mathbb{E} \begin{bmatrix} \dot{g}_i(\beta_1^\star) \\ W_{i,1}(\beta_1^\star) \dot{\ell}_i(\theta^\star) \end{bmatrix} \right)$$

Differentiability

$$= \sqrt{n} \left( \mathbb{P}_n \begin{bmatrix} \dot{g}_i(\hat{\beta}_1) \\ W_{i,1}(\hat{\beta}_1) \dot{\ell}_i(\hat{\theta}) \end{bmatrix} - \mathbb{E} \begin{bmatrix} \dot{g}_i(\hat{\beta}_1) \\ W_{i,1}(\hat{\beta}_1) \dot{\ell}_i(\hat{\theta}) \end{bmatrix} \right) + o_P(1)$$

$$\ddot{L} = \frac{\partial}{\partial \theta} \mathbb{E}_{\pi^\star} [\dot{\ell}_i(\theta)] \Big|_{\theta=\theta^\star}$$

$$= \sqrt{n} \left( \mathbb{E} \begin{bmatrix} \dot{g}_i(\beta_1^\star) \\ W_{i,1}(\beta_1^\star) \dot{\ell}_i(\theta^\star) \end{bmatrix} - \mathbb{E} \begin{bmatrix} \dot{g}_i(\hat{\beta}_1) \\ W_{i,1}(\hat{\beta}_1) \dot{\ell}_i(\hat{\theta}) \end{bmatrix} \right) + o_P(1)$$

$$= \begin{bmatrix} \ddot{G}_1 & 0 \\ V_1 & \ddot{L} \end{bmatrix}^{-1} \sqrt{n} \begin{pmatrix} \hat{\beta}_1 - \beta_1^\star \\ \hat{\theta} - \theta^\star \end{pmatrix} + \left( \begin{array}{c} \|\hat{\beta}_1 - \beta_1^\star\| \\ \|\hat{\theta} - \theta^\star\| \end{array} \right) + o_P(1)$$

$$\ddot{G}_1 = \frac{\partial}{\partial \beta} \mathbb{E} [\dot{g}_{1,i}(\beta)] \Big|_{\beta=\beta^\star}$$

$$V_1 = \frac{\partial}{\partial \beta_1} \mathbb{E}_{\pi_2(\beta_1)} [\dot{\ell}_i(\theta^\star)] \Big|_{\beta_1=\beta_1^\star}$$

# Differentiability

$$\frac{\partial}{\partial(\beta_1, \theta)} \mathbb{E} \begin{bmatrix} \dot{g}_i(\beta_1) \\ W_{i,1}(\beta_1) \dot{\ell}_i(\theta) \end{bmatrix} \Bigg|_{(\beta_1, \theta) = (\beta_1^*, \theta^*)}$$

$$= \mathbb{E} \begin{bmatrix} \frac{\partial}{\partial \beta_1} \dot{g}_i(\beta_1) \Big|_{\beta_1=\beta_1^*} & \frac{\partial}{\partial \theta} \dot{g}_i(\beta_1^*) \Big|_{\theta=\theta^*} \\ \frac{\partial}{\partial \beta_1} W_{i,1}(\beta_1) \Big|_{\beta_1=\beta_1^*} \quad \dot{\ell}_i(\theta^*) & W_{i,1}(\beta_1^*) \frac{\partial}{\partial \theta} \dot{\ell}_i(\theta) \Big|_{\theta=\theta^*} \end{bmatrix}$$

$$= \begin{bmatrix} \ddot{G}_1 & 0 \\ V_1 & \ddot{L} \end{bmatrix}$$

We've now outlined how to show step (2)

$$\begin{aligned}
 (2) \sqrt{n} & \left( \mathbb{P}_n \begin{bmatrix} \dot{g}_i(\beta_1^*) \\ W_{i,1}(\beta_1^*) \dot{\ell}_i(\theta^*) \end{bmatrix} - \mathbb{E} \begin{bmatrix} \dot{g}_i(\beta_1^*) \\ W_{i,1}(\beta_1^*) \dot{\ell}_i(\theta^*) \end{bmatrix} \right) \\
 &= \begin{bmatrix} \ddot{G}_1 & 0 \\ V_1 & \ddot{L} \end{bmatrix}^{-1} \sqrt{n} \begin{pmatrix} \hat{\beta}_1 - \beta_1^* \\ \hat{\theta} - \theta^* \end{pmatrix} + \left( \begin{array}{c} \|\hat{\beta}_1 - \beta_1^*\| \\ \|\hat{\theta} - \theta^*\| \end{array} \right) + o_P(1)
 \end{aligned}$$

## Proof Key Steps

$$\sqrt{n} \begin{pmatrix} \hat{\beta}_1 - \beta_1^* \\ \hat{\theta} - \theta^* \end{pmatrix} \xrightarrow{D} \mathcal{N}\left(0, \begin{bmatrix} \ddot{G}_1 & 0 \\ V_1 & \ddot{L} \end{bmatrix}^{-1} \Sigma_{1:2} \begin{bmatrix} \ddot{G}_1 & 0 \\ V_1 & \ddot{L} \end{bmatrix}^{-1, \top} \right)$$

$$(3) \sqrt{n} o_P\left(\begin{array}{c} \|\hat{\beta}_1 - \beta_1^*\| \\ \|\hat{\theta} - \theta^*\| \end{array}\right) = o_P(1)$$

HW Exercise (uses previous two steps)

# Joint Asymptotic Normality Result $T = 2$

$$\sqrt{n} \begin{pmatrix} \hat{\beta} - \beta^* \\ \hat{\theta} - \theta^* \end{pmatrix} \xrightarrow{D} \mathcal{N} \left( 0, \begin{bmatrix} \ddot{G}_1 & 0 \\ V_1 & \ddot{L} \end{bmatrix}^{-1} \Sigma_{1:2} \begin{bmatrix} \ddot{G}_1 & 0 \\ V_1 & \ddot{L} \end{bmatrix}^{-1, \top} \right)$$

Due to lower-triangular structure of “bread” matrices,

$$\sqrt{n} (\hat{\theta} - \theta^*) \xrightarrow{D} \mathcal{N} (0, \ddot{L}^{-1} \Sigma^{\text{adapt}} (\ddot{L}^{-1})^\top)$$

$$\Sigma^{\text{adapt}} = \mathbb{E}_{\pi^*} \left[ \left\{ \dot{\ell}_i(\theta^*) - V_1 \ddot{G}_1^{-1} \dot{g}_{i,1}(\beta_1^*) \right\} \left\{ \dot{\ell}_i(\theta^*) - V_1 \ddot{G}_1^{-1} \dot{g}_{i,1}(\beta_1^*) \right\}^\top \right]$$

# Slides for Reference

# Joint Asymptotic Normality Result (General $T$ )

$$\sqrt{n} \begin{pmatrix} \hat{\beta}_{1:T-1} - \beta_{1:T-1}^* \\ \hat{\theta} - \theta^* \end{pmatrix} \xrightarrow{D} \mathcal{N} \left( 0, \begin{bmatrix} \ddot{G}_{1:T-1} & 0 \\ V_{1:T-1} & \ddot{L} \end{bmatrix}^{-1} \Sigma_{1:T} \begin{bmatrix} \ddot{G}_{1:T-1} & 0 \\ V_{1:T-1} & \ddot{L} \end{bmatrix}^{-1,\top} \right)$$

$$\ddot{G}_{1:T-1} = \frac{\partial}{\partial \beta_{1:T-1}} \mathbb{E}_{\pi(\beta_{1:T-1})} \begin{bmatrix} \dot{g}_{1,i}(\beta_1) \\ \dot{g}_{2,i}(\beta_2) \\ \vdots \\ \dot{g}_{T-1,i}(\beta_{T-1}) \end{bmatrix} \Big|_{\beta_{1:T-1}=\beta_{1:T-1}^*}$$

$$V_{1:T-1} = \frac{\partial}{\partial \beta_{1:T-1}} \mathbb{E}_{\pi(\beta_{1:T-1})} [\dot{\ell}_i(\theta^*)] \Big|_{\beta_{1:T-1}=\beta_{1:T-1}^*}$$