Longitudinal weighted and trimmed treatment effects with flip interventions

Alec McClean, Iván Díaz, and Alex Levis

New York University Grossman School of Medicine

Motivation: positivity violations

Longitudinal data: **positivity assumption** says all subjects have non-zero prob. of all treatment regimes of interest (possibly 2^T regimes)

Positivity violations are typical with observational data

1. **"Absolute" violations:**
   zero propensity scores;
   estimand unidentified
2. **"Near"/"practical" violations:**
   near-zero propensity scores;
   estimators have large variance

Motivation: positivity violations

Longitudinal data: **positivity assumption** says all subjects have non-zero prob. of all treatment regimes of interest (possibly 2^T regimes)

Positivity violations are typical with observational data

1. **"Absolute" violations:**
   zero propensity scores;
   estimand unidentified
2. **"Near"/"practical" violations:**
   near-zero propensity scores;
   estimators have large variance

Motivation: positivity violations

Longitudinal data: **positivity assumption** says all subjects have non-zero prob. of all treatment regimes of interest (possibly 2^T regimes)

Positivity violations are typical with observational data

1. **"Absolute" violations:**
   zero propensity scores;
   estimand unidentified
2. **"Near"/"practical" violations:**
   near-zero propensity scores;
   estimators have large variance

Weighting/trimming with single-timepoint data

$\{(X_i, A_i, Y_i)\}_{i=1}^n \overset{iid}{\sim} \mathcal{P}$ with X_i covariates, $A_i \in \{0,1\}$ binary treatment and $Y_i \in \mathbb{R}$ outcome.

$Y(a)$ is potential outcome under treatment a.

Positivity violations: $\pi(X) = \mathbb{P}(A = 1 \mid X) \approx 0$ or ≈ 1.

Weighted average treatment effect (WATE):

Change population to satisfy positivity

$$\mathbb{E}\left[\frac{\mathbb{E}\{Y(1) - Y(0) \mid X\} f(X)}{\mathbb{E}\{f(X)\}}\right] \text{ with } \underbrace{f(X) : \mathcal{X} \to [0,1]}_{\text{often } s\{\pi(X)\}}$$

Weighting/trimming with single-timepoint data

$\{(X_i, A_i, Y_i)\}_{i=1}^n \overset{iid}{\sim} \mathcal{P}$ with X_i covariates, $A_i \in \{0, 1\}$ binary treatment and $Y_i \in \mathbb{R}$ outcome.
$Y(a)$ is potential outcome under treatment a.

Positivity violations: $\pi(X) = \mathbb{P}(A = 1 \mid X) \approx 0$ or ≈ 1.

Weighted average treatment effect (WATE):
Change population to satisfy positivity

$$\mathbb{E}\left[\frac{\mathbb{E}\{Y(1) - Y(0) \mid X\} f(X)}{\mathbb{E}\{f(X)\}}\right] \text{ with } \underbrace{f(X) : \mathcal{X} \to [0, 1]}_{\text{often } s\{\pi(X)\}}$$

Weighting/trimming with single-timepoint data

$\{(X_i, A_i, Y_i)\}_{i=1}^n \overset{iid}{\sim} \mathcal{P}$ with X_i covariates, $A_i \in \{0,1\}$ binary treatment and $Y_i \in \mathbb{R}$ outcome.
$Y(a)$ is potential outcome under treatment a.

Positivity violations: $\pi(X) = \mathbb{P}(A = 1 \mid X) \approx 0$ or ≈ 1.

Weighted average treatment effect (WATE):
Change population to satisfy positivity

$$\mathbb{E}\left[\frac{\mathbb{E}\{Y(1) - Y(0) \mid X\} f(X)}{\mathbb{E}\{f(X)\}}\right] \text{ with } \underbrace{f(X) : \mathcal{X} \to [0,1]}_{\text{often } s\{\pi(X)\}}$$

Weighting/trimming with single-timepoint data

$\{(X_i, A_i, Y_i)\}_{i=1}^n \overset{iid}{\sim} \mathcal{P}$ with X_i covariates, $A_i \in \{0, 1\}$ binary treatment and $Y_i \in \mathbb{R}$ outcome.
$Y(a)$ is potential outcome under treatment a.

Positivity violations: $\pi(X) = \mathbb{P}(A = 1 \mid X) \approx 0$ or ≈ 1.

Weighted average treatment effect (WATE):
Change population to satisfy positivity

$$\mathbb{E}\left[\frac{\mathbb{E}\{Y(1) - Y(0) \mid X\}f(X)}{\mathbb{E}\{f(X)\}}\right] \text{ with } \underbrace{f(X) : \mathcal{X} \to [0, 1]}_{\text{often } s\{\pi(X)\}}$$

Example weights

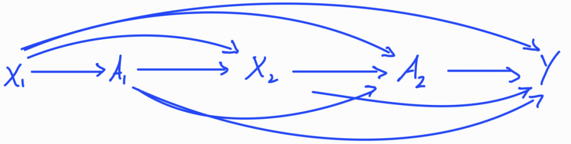
WATE	Weight; $f(X)$
ATO	$\pi(X)(1-\pi(X))$
Trimmed ATE	$\mathbb{1}\{\varepsilon \le \pi(X) \le 1-\varepsilon\}$ for $\varepsilon \ge 0$
Smooth trimmed ATE	$S\{\pi(X);\varepsilon\}$ where $S(x;\varepsilon)$ approximates $\mathbb{1}(\varepsilon \le x \le 1-\varepsilon)$
ATE	1
ATT	$\pi(X)$
ATC	$1-\pi(X)$
Matching-style ATE	$\pi(X) \wedge 1-\pi(X)$
Direct covariate balancing	Varies and is data-dependent; derived to directly balance covariates

Example weights

WATE	**Weight;** $f(X)$
ATO	$\pi(X)(1-\pi(X))$
Trimmed ATE	$\mathbb{1}\{\varepsilon \leq \pi(X) \leq 1-\varepsilon\}$ for $\varepsilon \geq 0$
Smooth trimmed ATE	$S\{\pi(X);\varepsilon\}$ where $S(x;\varepsilon)$ approximates $\mathbb{1}(\varepsilon \leq x \leq 1-\varepsilon)$
ATE	1
ATT	$\pi(X)$
ATC	$1-\pi(X)$
Matching-style ATE	$\pi(X) \wedge 1-\pi(X)$
Direct covariate balancing	Varies and is data-dependent; derived to directly balance covariates

Weighting on non-baseline covariates is challenging

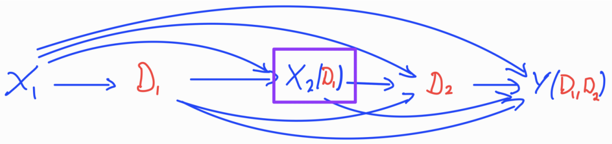
Data: (X_1, A_1, X_2, A_2, Y).



How to weight second timepoint propensity scores?

(where we often have positivity violations!)

Traditional answer: not possible

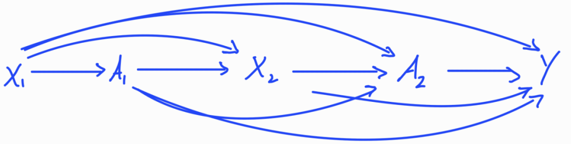


1. After intervention D_1, $X_2(D_1)$ can be counterfactual.
2. Therefore, $X_2(D_1)$ not necessarily equal to X_2

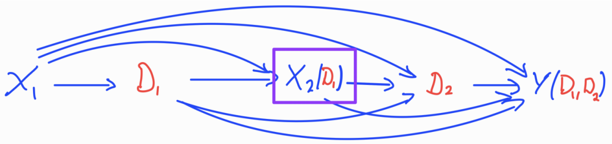
$\Longrightarrow$ Weighting/trimming on X_2 is a bad idea. [Petersen et al., 2012]

Weighting on non-baseline covariates is challenging

Data: (X_1, A_1, X_2, A_2, Y).



How to weight second timepoint propensity scores?
(where we often have positivity violations!)
Traditional answer: not possible

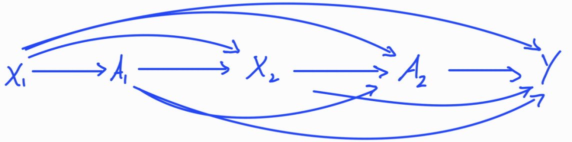


1. After intervention D_1, $X_2(D_1)$ can be counterfactual.
2. Therefore, $X_2(D_1)$ not necessarily equal to X_2

$\Longrightarrow$ Weighting/trimming on X_2 is a bad idea. [Petersen et al., 2012]

Weighting on non-baseline covariates is challenging

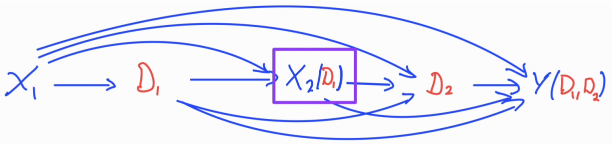
Data: (X_1, A_1, X_2, A_2, Y).



How to weight second timepoint propensity scores?

(where we often have positivity violations!)

Traditional answer: not possible

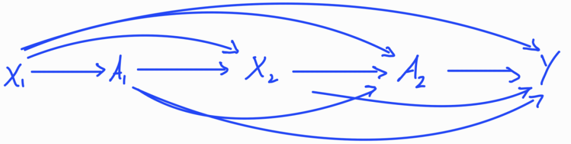


1. After intervention D_1, $X_2(D_1)$ can be counterfactual.
2. Therefore, $X_2(D_1)$ not necessarily equal to X_2

$\Longrightarrow$ Weighting/trimming on X_2 is a bad idea. [Petersen et al., 2012]

Weighting on non-baseline covariates is challenging

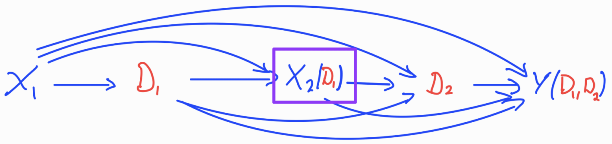
Data: (X_1, A_1, X_2, A_2, Y).



How to weight second timepoint propensity scores?

(where we often have positivity violations!)

Traditional answer: not possible

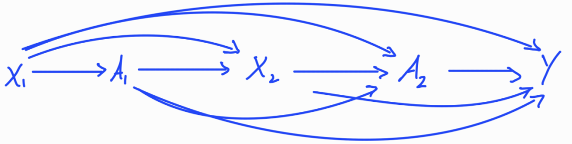


1. After intervention D_1, $X_2(D_1)$ can be counterfactual.
2. Therefore, $X_2(D_1)$ not necessarily equal to X_2

$\Longrightarrow$ Weighting/trimming on X_2 is a bad idea. [Petersen et al., 2012]

Weighting on non-baseline covariates is challenging

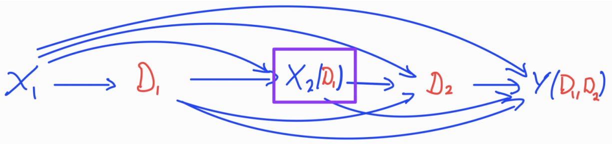
Data: (X_1, A_1, X_2, A_2, Y).



How to weight second timepoint propensity scores?

(where we often have positivity violations!)

Traditional answer: not possible



1. After intervention D_1, $X_2(D_1)$ can be counterfactual.
2. Therefore, $X_2(D_1)$ not necessarily equal to X_2

$\Longrightarrow$ Weighting/trimming on X_2 is a bad idea. [Petersen et al., 2012]

Dynamic stochastic interventions

$D = d(A, X, V)$ can depend on treatment A, covariates X, and auxiliary randomness V

D can adapt so $\mathbb{E}\{Y(D)\}$ IDable and estimable

- Incremental propensity score interventions [Kennedy, 2019]
  $D \sim \text{Bernoulli}\{Q_\delta(X)\}$,
  $\text{odds}\{Q_\delta(X)\} = \delta \cdot \text{odds}\{\pi(X)\}$
- Modified treatment policies [Haneuse and Rotnitzky, 2013, Díaz et al., 2023]
  $D = A + \delta$

Dynamic stochastic interventions

$D = d(A, X, V)$ can depend on treatment A, covariates X, and auxiliary randomness V

D can adapt so $\mathbb{E}\{Y(D)\}$ IDable and estimable

- Incremental propensity score interventions [Kennedy, 2019]
  $D \sim \text{Bernoulli}\{Q_\delta(X)\}$,
  $\text{odds}\{Q_\delta(X)\} = \delta \cdot \text{odds}\{\pi(X)\}$
- Modified treatment policies [Haneuse and Rotnitzky, 2013, Díaz et al., 2023]
  $D = A + \delta$

Dynamic stochastic interventions

$D = d(A, X, V)$ can depend on treatment A, covariates X, and auxiliary randomness V

D can adapt so $\mathbb{E}\{Y(D)\}$ IDable and estimable

- Incremental propensity score interventions [Kennedy, 2019]
  $D \sim \text{Bernoulli}\{Q_\delta(X)\}$,
  $\text{odds}\{Q_\delta(X)\} = \delta \cdot \text{odds}\{\pi(X)\}$
- Modified treatment policies [Haneuse and Rotnitzky, 2013, Díaz et al., 2023]
  $D = A + \delta$

Dynamic stochastic interventions

$D = d(A, X, V)$ can depend on treatment A, covariates X, and auxiliary randomness V

D can adapt so $\mathbb{E}\{Y(D)\}$ IDable and estimable

- Incremental propensity score interventions [Kennedy, 2019]
  $D \sim \text{Bernoulli}\{Q_\delta(X)\}$,
  $\text{odds}\{Q_\delta(X)\} = \delta \cdot \text{odds}\{\pi(X)\}$
- Modified treatment policies [Haneuse and Rotnitzky, 2013, Díaz et al., 2023]
  $D = A + \delta$

Pairs of flip interventions in one timepoint

Flip intervention: Given $f(X)$, define a pair of flip interventions, one for each $a \in \{0, 1\}$, as

$$\boxed{D_f(a) = \begin{cases} A & \text{if } A = a, \\ a\mathbb{1}\{V \leq f(X)\} + A\mathbb{1}\{V > f(X)\} & \text{otherwise,} \end{cases}}$$

where $V \sim \text{Unif}(0, 1)$.

If $A = a$, **do not intervene**;
o/w, **flip** to a **with probability $f(X)$.**

WATEs are *interventional* flip effects

If $A = a$, **do not intervene**; o/w, **flip** to a **with probability** $f(X)$.

***Interventional* flip effect**: given $D_f(1), D_f(0)$,

$$\psi_f := \frac{\mathbb{E}\left[Y\{D_f(1)\} - Y\{D_f(0)\}\right]}{\mathbb{E}\left\{D_f(1) - D_f(0)\right\}}$$

This is the average effect on potential outcomes of $D_f(1)$ compared to $D_f(0)$ *per unit of additional treatment.* [Zhou and Opacic, 2022]

Proposition: Interventional flip effects are WATEs:

$$\psi_f = \mathbb{E}\left[\frac{\mathbb{E}\{Y(1) - Y(0) \mid X\}f(X)}{\mathbb{E}\{f(X)\}}\right].$$

Move the goalposts: we can define and estimate longitudinal interventional flip effects

WATEs are *interventional* flip effects

If $A = a$, **do not intervene**; o/w, **flip** to a **with probability** $f(X)$**.**

***Interventional* flip effect**: given $D_f(1), D_f(0)$,

$$\psi_f := \frac{\mathbb{E}\left[Y\{D_f(1)\} - Y\{D_f(0)\}\right]}{\mathbb{E}\left\{D_f(1) - D_f(0)\right\}}$$

This is the average effect on potential outcomes of $D_f(1)$ compared to $D_f(0)$ *per unit of additional treatment.* [Zhou and Opacic, 2022]

Proposition: Interventional flip effects are WATEs:

$$\psi_f = \mathbb{E}\left[\frac{\mathbb{E}\{Y(1) - Y(0) \mid X\}f(X)}{\mathbb{E}\{f(X)\}}\right].$$

Move the goalposts: we can define and estimate longitudinal interventional flip effects

WATEs are *interventional* flip effects

If $A = a$, **do not intervene**; o/w, **flip** to a **with probability** $f(X)$.

***Interventional* flip effect**: given $D_f(1), D_f(0)$,

$$\psi_f := \frac{\mathbb{E}\left[Y\{D_f(1)\} - Y\{D_f(0)\}\right]}{\mathbb{E}\left\{D_f(1) - D_f(0)\right\}}$$

This is the average effect on potential outcomes of $D_f(1)$ compared to $D_f(0)$ *per unit of additional treatment.* [Zhou and Opacic, 2022]

Proposition: Interventional flip effects are WATEs:

$$\psi_f = \mathbb{E}\left[\frac{\mathbb{E}\{Y(1) - Y(0) \mid X\} f(X)}{\mathbb{E}\{f(X)\}}\right].$$

Move the goalposts: we can define and estimate longitudinal interventional flip effects

WATEs are *interventional* flip effects

If $A = a$, **do not intervene**; o/w, **flip** to a **with probability** $f(X)$.

***Interventional* flip effect**: given $D_f(1), D_f(0)$,

$$\psi_f := \frac{\mathbb{E}\left[Y\{D_f(1)\} - Y\{D_f(0)\}\right]}{\mathbb{E}\left\{D_f(1) - D_f(0)\right\}}$$

This is the average effect on potential outcomes of $D_f(1)$ compared to $D_f(0)$ *per unit of additional treatment.* [Zhou and Opacic, 2022]

Proposition: Interventional flip effects are WATEs:

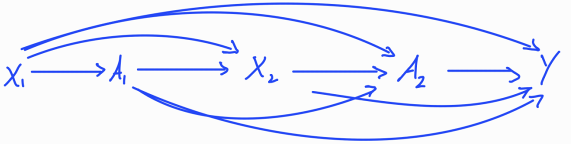
$$\psi_f = \mathbb{E}\left[\frac{\mathbb{E}\{Y(1) - Y(0) \mid X\} f(X)}{\mathbb{E}\{f(X)\}}\right].$$

Move the goalposts: we can define and estimate longitudinal interventional flip effects

Longitudinal setup

Longitudinal data

$$\{Z_i\}_{i=1}^n \overset{iid}{\sim} \mathbb{P} \in \mathcal{P} \text{ where } Z = (X_1, A_1, X_2, A_2, \ldots, X_T, A_T, Y)$$



$X_t \in \mathbb{R}^d$: time-varying covariates
$A_t \in \{0, 1\}$: time-varying binary treatment
$Y \in \mathbb{R}$: ultimate outcome of interest

History of O_t at t: $\overline{O}_t = (O_1, \ldots, O_t)$
Future of O_t from t: $\underline{O}_t = (O_t, \ldots, O_T)$
$H_t = (\overline{X}_t, \overline{A}_{t-1})$: covariate and treatment history at t

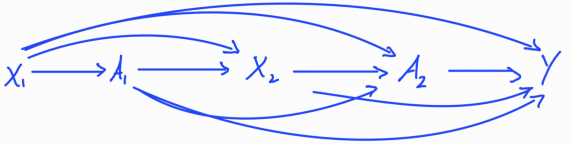
NPSEM assumption: there are $\{f_{X,t}, f_{A,t}\}_{t=1}^T$ and f_Y such that

$$X_t = f_{X,t}(A_{t-1}, H_{t-1}, U_{X,t}),$$
$$A_t = f_{A,t}(H_t, U_{A,t}), \text{ and}$$
$$Y = f_Y(A_T, H_T, U_Y).$$

where $\{U_{X,t}, U_{A,t}, U_Y\}$ are exogeneous variables

Longitudinal data

$\{Z_i\}_{i=1}^n \overset{iid}{\sim} \mathbb{P} \in \mathcal{P}$ where $Z = (X_1, A_1, X_2, A_2, \ldots, X_T, A_T, Y)$



$X_t \in \mathbb{R}^d$: time-varying covariates
$A_t \in \{0, 1\}$: time-varying binary treatment
$Y \in \mathbb{R}$: ultimate outcome of interest

History of O_t at t: $\overline{O}_t = (O_1, \ldots, O_t)$
Future of O_t from t: $\underline{O}_t = (O_t, \ldots, O_T)$
$H_t = (\overline{X}_t, \overline{A}_{t-1})$: covariate and treatment history at t

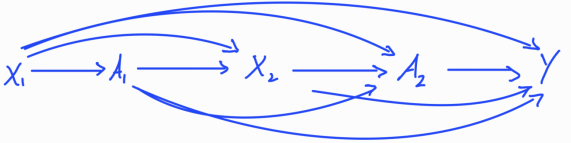
NPSEM assumption: there are $\{f_{X,t}, f_{A,t}\}_{t=1}^T$ and f_Y such that

$$X_t = f_{X,t}(A_{t-1}, H_{t-1}, U_{X,t}),$$
$$A_t = f_{A,t}(H_t, U_{A,t}), \text{ and}$$
$$Y = f_Y(A_T, H_T, U_Y).$$

where $\{U_{X,t}, U_{A,t}, U_Y\}$ are exogeneous variables

Longitudinal data

$\{Z_i\}_{i=1}^n \overset{iid}{\sim} \mathbb{P} \in \mathcal{P}$ where $Z = (X_1, A_1, X_2, A_2, \ldots, X_T, A_T, Y)$



$X_t \in \mathbb{R}^d$: time-varying covariates
$A_t \in \{0, 1\}$: time-varying binary treatment
$Y \in \mathbb{R}$: ultimate outcome of interest

History of O_t at t: $\overline{O}_t = (O_1, \ldots, O_t)$
Future of O_t from t: $\underline{O}_t = (O_t, \ldots, O_T)$
$H_t = (\overline{X}_t, \overline{A}_{t-1})$: covariate and treatment history at t

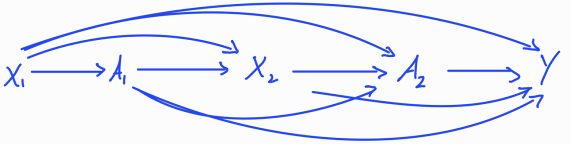
NPSEM assumption: there are $\{f_{X,t}, f_{A,t}\}_{t=1}^T$ and f_Y such that

$$X_t = f_{X,t}(A_{t-1}, H_{t-1}, U_{X,t}),$$
$$A_t = f_{A,t}(H_t, U_{A,t}), \text{ and}$$
$$Y = f_Y(A_T, H_T, U_Y).$$

where $\{U_{X,t}, U_{A,t}, U_Y\}$ are exogeneous variables

Longitudinal data

$\{Z_i\}_{i=1}^n \overset{iid}{\sim} \mathbb{P} \in \mathcal{P}$ where $Z = (X_1, A_1, X_2, A_2, \ldots, X_T, A_T, Y)$



$X_t \in \mathbb{R}^d$: time-varying covariates
$A_t \in \{0, 1\}$: time-varying binary treatment
$Y \in \mathbb{R}$: ultimate outcome of interest

History of O_t at t: $\overline{O}_t = (O_1, \ldots, O_t)$
Future of O_t from t: $\underline{O}_t = (O_t, \ldots, O_T)$
$H_t = (\overline{X}_t, \overline{A}_{t-1})$: covariate and treatment history at t

NPSEM assumption: there are $\{f_{X,t}, f_{A,t}\}_{t=1}^T$ and f_Y such that

$$X_t = f_{X,t}(A_{t-1}, H_{t-1}, U_{X,t}),$$
$$A_t = f_{A,t}(H_t, U_{A,t}), \text{ and}$$
$$Y = f_Y(A_T, H_T, U_Y).$$

where $\{U_{X,t}, U_{A,t}, U_Y\}$ are exogeneous variables

Causal assumptions

NPSEM embeds **consistency**
$\overline{D}_T = \overline{a}_T \implies Y(\overline{D}_T) = Y(\overline{a}_T)$

We will avoid **positivity** by weighting/flipping
(e.g. $\exists\, \delta > 0$ s.t. $\mathbb{P}\{\delta < \mathbb{P}(A_t = 1 \mid H_t) < 1 - \delta\} = 1$.)

Strong sequential randomization:

$$U_{A,t} \perp\!\!\!\perp (\underline{U}_{X,t+1}, \underline{U}_{A,t+1}, U_Y) \mid H_t$$

Common causes of A_t and future covariates and outcome **and treatments** measured. Allows for ID when intervention depends on natural value of treatment.

We can construct interventions that only need *standard* SR for ID.
$U_{A,t} \perp\!\!\!\perp (\underline{U}_{X,t+1}, U_Y) \mid H_t$ for all $t \leq T$
Common causes of A_t and future covariates and outcome measured.

Causal assumptions

NPSEM embeds **consistency**
$\overline{D}_T = \overline{a}_T \implies Y(\overline{D}_T) = Y(\overline{a}_T)$

We will avoid **positivity** by weighting/flipping
(e.g. $\exists\, \delta > 0$ s.t. $\mathbb{P}\{\delta < \mathbb{P}(A_t = 1 \mid H_t) < 1 - \delta\} = 1$.)

Strong sequential randomization:

$$U_{A,t} \perp\!\!\!\perp (\underline{U}_{X,t+1}, \underline{U}_{A,t+1}, U_Y) \mid H_t$$

Common causes of A_t and future covariates and outcome **and treatments** measured. Allows for ID when intervention depends on natural value of treatment.

We can construct interventions that only need *standard* SR for ID.
$U_{A,t} \perp\!\!\!\perp (\underline{U}_{X,t+1}, U_Y) \mid H_t$ for all $t \leq T$
Common causes of A_t and future covariates and outcome measured.

Causal assumptions

NPSEM embeds **consistency**
$\overline{D}_T = \overline{a}_T \implies Y(\overline{D}_T) = Y(\overline{a}_T)$

We will avoid **positivity** by weighting/flipping
(e.g. $\exists\, \delta > 0$ s.t. $\mathbb{P}\{\delta < \mathbb{P}(A_t = 1 \mid H_t) < 1 - \delta\} = 1$.)

Strong sequential randomization:

$$U_{A,t} \perp\!\!\!\perp (\underline{U}_{X,t+1}, \underline{U}_{A,t+1}, U_Y) \mid H_t$$

Common causes of A_t and future covariates and outcome **and treatments** measured. Allows for ID when intervention depends on natural value of treatment.

We can construct interventions that only need *standard* SR for ID.
$U_{A,t} \perp\!\!\!\perp (\underline{U}_{X,t+1}, U_Y) \mid H_t$ for all $t \leq T$
Common causes of A_t and future covariates and outcome measured.

Causal assumptions

NPSEM embeds **consistency**

$\overline{D}_T = \overline{a}_T \implies Y(\overline{D}_T) = Y(\overline{a}_T)$

We will avoid **positivity** by weighting/flipping

(e.g. $\exists\, \delta > 0$ s.t. $\mathbb{P}\{\delta < \mathbb{P}(A_t = 1 \mid H_t) < 1 - \delta\} = 1$.)

Strong sequential randomization:

$$U_{A,t} \perp\!\!\!\perp (\underline{U}_{X,t+1}, \underline{U}_{A,t+1}, U_Y) \mid H_t$$

Common causes of A_t and future covariates and outcome **and treatments** measured. Allows for ID when intervention depends on natural value of treatment.

We can construct interventions that only need *standard* SR for ID.

$U_{A,t} \perp\!\!\!\perp (\underline{U}_{X,t+1}, U_Y) \mid H_t$ for all $t \leq T$

Common causes of A_t and future covariates and outcome measured.

Causal assumptions

NPSEM embeds **consistency**
$\overline{D}_T = \overline{a}_T \implies Y(\overline{D}_T) = Y(\overline{a}_T)$

We will avoid **positivity** by weighting/flipping
(e.g. $\exists\ \delta > 0$ s.t. $\mathbb{P}\{\delta < \mathbb{P}(A_t = 1 \mid H_t) < 1 - \delta\} = 1$.)

Strong sequential randomization:

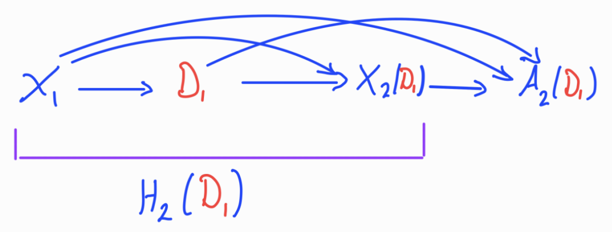
$$U_{A,t} \perp\!\!\!\perp (\underline{U}_{X,t+1}, \underline{U}_{A,t+1}, U_Y) \mid H_t$$

Common causes of A_t and future covariates and outcome **and treatments** measured. Allows for ID when intervention depends on natural value of treatment.

We can construct interventions that only need *standard* SR for ID.
$U_{A,t} \perp\!\!\!\perp (\underline{U}_{X,t+1}, U_Y) \mid H_t$ for all $t \leq T$
Common causes of A_t and future covariates and outcome measured.

$\overline{D}_{t-1}$ generates **counterfactual variables at time t**:

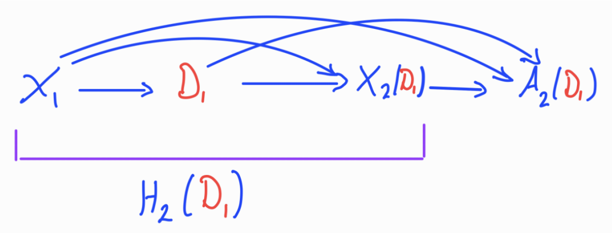
- $X_t(\overline{D}_{t-1}) = f_{X,t}(D_{t-1}, H_{t-1}(\overline{D}_{t-2}), U_{X,t})$
  "Natural covariate value"
- $H_t(\overline{D}_{t-1}) = (\overline{D}_{t-1}, \overline{X}_t(\overline{D}_{t-1}))$
  "Natural covariate and treatment history"
- $A_t(\overline{D}_{t-1}) = f_{A,t}(H_t(\overline{D}_{t-1}), U_{A,t})$
  "Natural value of treatment"



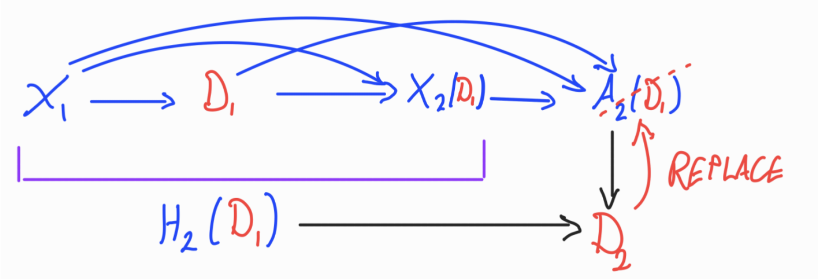
Counterfactual variables under $\overline{D}_{t-1}$

$\overline{D}_{t-1}$ generates **counterfactual variables at time** t:

- $X_t(\overline{D}_{t-1}) = f_{X,t}(D_{t-1}, H_{t-1}(\overline{D}_{t-2}), U_{X,t})$
  "Natural covariate value"
- $H_t(\overline{D}_{t-1}) = (\overline{D}_{t-1}, \overline{X}_t(\overline{D}_{t-1}))$
  "Natural covariate and treatment history"
- $A_t(\overline{D}_{t-1}) = f_{A,t}(H_t(\overline{D}_{t-1}), U_{A,t})$
  "Natural value of treatment"



Intervention D_t and potential outcomes $\overline{D}_T$

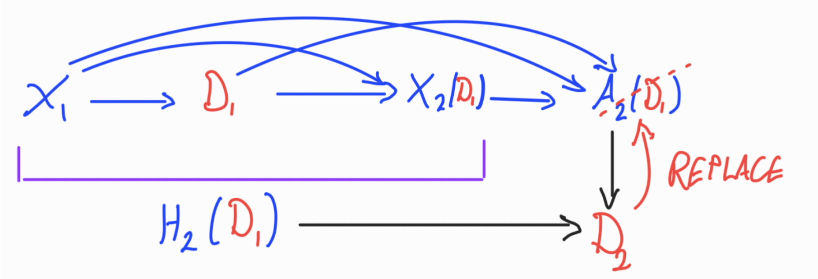


D_t can be a function of $A_t(\overline{D}_{t-1}), H_t(\overline{D}_{t-1})$.

Ultimately, replace $\overline{A}_T$ with $\overline{D}_T$:

- $Y(\overline{D}_T) = f_Y(D_T, H_T(\overline{D}_{T-1}), U_Y)$
  "Counterfactual outcomes under $\overline{D}_T$"

Intervention D_t and potential outcomes $\overline{D}_T$



D_t can be a function of $A_t(\overline{D}_{t-1}), H_t(\overline{D}_{t-1})$.

Ultimately, replace $\overline{A}_T$ with $\overline{D}_T$:

- $Y(\overline{D}_T) = f_Y(D_T, H_T(\overline{D}_{T-1}), U_Y)$
  "Counterfactual outcomes under $\overline{D}_T$"

Longitudinal weighting and trimming with flip interventions

Longitudinal weights

Let $\overline{a}_T \in \{0,1\}^T$ be the target regime

$f_t^c(a_t; h_t) \ : \ \{0,1\} \times \mathcal{H}_t \ \rightarrow \ [0,1]$ is a ***counterfactual weight function***

Examples: $\pi_t^c\{H_t(\overline{D}_{t-1})\} := \mathbb{P}\{A_t(\overline{D}_{t-1}) = a_t \mid H_t(\overline{D}_{t-1})\}$

Type of weighting	**Flipping prob.;** $f_t^c\{a_t; H_t(\overline{D}_{t-1})\}$
ATT-style	$\pi_t^c\{H_t(\overline{D}_{t-1})\}$
Overlap weighting	$\pi_t^c\{H_t(\overline{D}_{t-1})\}\left[1 - \pi_t^c\{H_t(\overline{D}_{t-1})\}\right]$
Trimming	$\mathbb{1}\{\varepsilon \leq \pi_t^c\{H_t(\overline{D}_{t-1})\} \leq 1 - \varepsilon\}$
Smooth trimming	$S\left[\pi_t^c\{H_t(\overline{D}_{t-1})\}; \varepsilon\right]$ where $S(x;\varepsilon)$ approximates $\mathbb{1}(\varepsilon < x < 1 - \varepsilon)$

(Just replace π from single-timpoint with π_t^c)

Longitudinal weights

Let $\overline{a}_T \in \{0,1\}^T$ be the target regime

$f_t^c(a_t; h_t) \ : \ \{0,1\} \times \mathcal{H}_t \ \to \ [0,1]$ is a ***counterfactual* weight function**

Examples: $\pi_t^c\{H_t(\overline{D}_{t-1})\} := \mathbb{P}\{A_t(\overline{D}_{t-1}) = a_t \mid H_t(\overline{D}_{t-1})\}$

Type of weighting	Flipping prob.; $f_t^c\{a_t; H_t(\overline{D}_{t-1})\}$
ATT-style	$\pi_t^c\{H_t(\overline{D}_{t-1})\}$
Overlap weighting	$\pi_t^c\{H_t(\overline{D}_{t-1})\}\left[1 - \pi_t^c\{H_t(\overline{D}_{t-1})\}\right]$
Trimming	$\mathbb{1}\{\varepsilon \leq \pi_t^c\{H_t(\overline{D}_{t-1})\} \leq 1 - \varepsilon\}$
Smooth trimming	$S\left[\pi_t^c\{H_t(\overline{D}_{t-1})\}; \varepsilon\right]$ where $S(x;\varepsilon)$ approximates $\mathbb{1}(\varepsilon < x < 1-\varepsilon)$

(Just replace π from single-timpoint with π_t^c)

Longitudinal weights

Let $\overline{a}_T \in \{0,1\}^T$ be the target regime

$f_t^c(a_t; h_t) \; : \; \{0,1\} \times \mathcal{H}_t \; \to \; [0,1]$ is a ***counterfactual weight function***

Examples: $\pi_t^c\{H_t(\overline{D}_{t-1})\} := \mathbb{P}\{A_t(\overline{D}_{t-1}) = a_t \mid H_t(\overline{D}_{t-1})\}$

Type of weighting	**Flipping prob.;** $f_t^c\{a_t; H_t(\overline{D}_{t-1})\}$
ATT-style	$\pi_t^c\{H_t(\overline{D}_{t-1})\}$
Overlap weighting	$\pi_t^c\{H_t(\overline{D}_{t-1})\}\left[1 - \pi_t^c\{H_t(\overline{D}_{t-1})\}\right]$
Trimming	$\mathbb{1}\{\varepsilon \leq \pi_t^c\{H_t(\overline{D}_{t-1})\} \leq 1 - \varepsilon\}$
Smooth trimming	$S\left[\pi_t^c\{H_t(\overline{D}_{t-1})\}; \varepsilon\right]$ where $S(x;\varepsilon)$ approximates $\mathbb{1}(\varepsilon < x < 1 - \varepsilon)$

(Just replace π from single-timpoint with π_t^c)

Longitudinal flip interventions

Given a weight function $f_t^c \equiv f_t^c\{a_t; H_t(\overline{D}_{t-1})\}$

Flip intervention at time t targeting a_t is

$$\boxed{D_t = \begin{cases} A_t(\overline{D}_{t-1}), & \text{if } A_t(\overline{D}_{t-1}) = a_t, \\ \mathbb{1}\left(V_t \leq f_t^c\right) a_t + \mathbb{1}\left(V_t > f_t^c\right) A_t(\overline{D}_{t-1}), & \text{otherwise,} \end{cases}}$$

where $\{V_1, \ldots, V_T\} \overset{iid}{\sim} \text{Unif}(0,1)$ and $\{V_1, \ldots, V_T\} \perp\!\!\!\perp Z$.

If $A_t(\overline{D}_{t-1}) = a_t$, **do nothing**;
otherwise, **flip** to a_t **with probability** $f_t^c\{a_t; H_t(\overline{D}_{t-1})\}$

Longitudinal flip interventions

Given a weight function $f_t^c \equiv f_t^c\{a_t; H_t(\overline{D}_{t-1})\}$

Flip intervention at time t targeting a_t is

$$\boxed{D_t = \begin{cases} A_t(\overline{D}_{t-1}), & \text{if } A_t(\overline{D}_{t-1}) = a_t, \\ \mathbb{1}\left(V_t \le f_t^c\right) a_t + \mathbb{1}\left(V_t > f_t^c\right) A_t(\overline{D}_{t-1}), & \text{otherwise,} \end{cases}}$$

where $\{V_1, \ldots, V_T\} \overset{iid}{\sim} \text{Unif}(0,1)$ and $\{V_1, \ldots, V_T\} \perp\!\!\!\perp Z$.

If $A_t(\overline{D}_{t-1}) = a_t$, **do nothing**;
otherwise, **flip** to a_t **with probability** $f_t^c\{a_t; H_t(\overline{D}_{t-1})\}$

Longitudinal *interventional* flip effects

$\overline{D}_T$ targets $\overline{a}_T$; $\overline{D}'_T$ targets $\overline{a}'_T$

Longitudinal interventional flip effect:

$$\boxed{\frac{\mathbb{E}\left\{Y(\overline{D}_T) - Y(\overline{D}'_T)\right\}}{\frac{1}{T}\sum_{t=1}^{T}\left|\mathbb{E}\left(D_t - D'_t\right)\right|}}$$

- $\mathbb{E}\left\{Y(\overline{D}_T) - Y(\overline{D}'_T)\right\}$: same as before
- $T^{-1}\sum_{t=1}^{T}\left|\mathbb{E}\left(D_t - D'_t\right)\right|$: *average absolute per-timepoint change in the number of treatments*

There are multiple options for the denominator

1. $T^{-1}\sum_{t=1}^{T}\mathbb{P}\left(D_t \neq D'_t\right)$: *Average per-timepoint probability of switching treatment*
2. *Joint distributional distance* between dist. of $\overline{D}_T$ and dist. of $\overline{D}'_T$ (e.g., f-divergence, optimal transport metric).

Longitudinal *interventional* flip effects

$\overline{D}_T$ targets $\overline{a}_T$; $\overline{D}'_T$ targets $\overline{a}'_T$

Longitudinal interventional flip effect:

$$\boxed{\frac{\mathbb{E}\left\{Y(\overline{D}_T) - Y(\overline{D}'_T)\right\}}{\frac{1}{T}\sum_{t=1}^{T}\left|\mathbb{E}\left(D_t - D'_t\right)\right|}}$$

- $\mathbb{E}\left\{Y(\overline{D}_T) - Y(\overline{D}'_T)\right\}$: same as before
- $T^{-1}\sum_{t=1}^{T}\left|\mathbb{E}\left(D_t - D'_t\right)\right|$: *average absolute per-timepoint change in the number of treatments*

There are multiple options for the denominator

1. $T^{-1}\sum_{t=1}^{T}\mathbb{P}\left(D_t \neq D'_t\right)$: *Average per-timepoint probability of switching treatment*
2. *Joint distributional distance* between dist. of $\overline{D}_T$ and dist. of $\overline{D}'_T$ (e.g., f-divergence, optimal transport metric).

Longitudinal *interventional* flip effects

$\overline{D}_T$ targets $\overline{a}_T$; $\overline{D}'_T$ targets $\overline{a}'_T$

Longitudinal interventional flip effect:

$$\boxed{\frac{\mathbb{E}\left\{Y(\overline{D}_T) - Y(\overline{D}'_T)\right\}}{\frac{1}{T}\sum_{t=1}^{T} \left|\mathbb{E}\left(D_t - D'_t\right)\right|}}$$

- $\mathbb{E}\left\{Y(\overline{D}_T) - Y(\overline{D}'_T)\right\}$: same as before
- $T^{-1}\sum_{t=1}^{T} \left|\mathbb{E}\left(D_t - D'_t\right)\right|$: *average absolute per-timepoint change in the number of treatments*

There are multiple options for the denominator

1. $T^{-1}\sum_{t=1}^{T} \mathbb{P}\left(D_t \neq D'_t\right)$: *Average per-timepoint probability of switching treatment*
2. *Joint distributional distance* between dist. of $\overline{D}_T$ and dist. of $\overline{D}'_T$ (e.g., f-divergence, optimal transport metric).

Longitudinal *interventional* flip effects

$\overline{D}_T$ targets $\overline{a}_T$; $\overline{D}'_T$ targets $\overline{a}'_T$

Longitudinal interventional flip effect:

$$\boxed{\frac{\mathbb{E}\left\{Y(\overline{D}_T) - Y(\overline{D}'_T)\right\}}{\frac{1}{T}\sum_{t=1}^{T}\left|\mathbb{E}\left(D_t - D'_t\right)\right|}}$$

- $\mathbb{E}\left\{Y(\overline{D}_T) - Y(\overline{D}'_T)\right\}$: same as before
- $T^{-1}\sum_{t=1}^{T}\left|\mathbb{E}\left(D_t - D'_t\right)\right|$: *average absolute per-timepoint change in the number of treatments*

There are multiple options for the denominator

1. $T^{-1}\sum_{t=1}^{T}\mathbb{P}\left(D_t \neq D'_t\right)$: *Average per-timepoint probability of switching treatment*
2. *Joint distributional distance* between dist. of $\overline{D}_T$ and dist. of $\overline{D}'_T$ (e.g., f-divergence, optimal transport metric).

Longitudinal *interventional* flip effects

$\overline{D}_T$ targets $\overline{a}_T$; $\overline{D}'_T$ targets $\overline{a}'_T$

Longitudinal interventional flip effect:

$$\boxed{\frac{\mathbb{E}\left\{Y(\overline{D}_T) - Y(\overline{D}'_T)\right\}}{\frac{1}{T}\sum_{t=1}^{T}\left|\mathbb{E}\left(D_t - D'_t\right)\right|}}$$

- $\mathbb{E}\left\{Y(\overline{D}_T) - Y(\overline{D}'_T)\right\}$: same as before
- $T^{-1}\sum_{t=1}^{T}\left|\mathbb{E}\left(D_t - D'_t\right)\right|$: *average absolute per-timepoint change in the number of treatments*

There are multiple options for the denominator

1. $T^{-1}\sum_{t=1}^{T}\mathbb{P}\left(D_t \neq D'_t\right)$: *Average per-timepoint probability of switching treatment*
2. *Joint distributional distance* between dist. of $\overline{D}_T$ and dist. of $\overline{D}'_T$ (e.g., f-divergence, optimal transport metric).

Longitudinal *interventional* flip effects

$\overline{D}_T$ targets $\overline{a}_T$; $\overline{D}'_T$ targets $\overline{a}'_T$

Longitudinal interventional flip effect:

$$\boxed{\frac{\mathbb{E}\left\{Y(\overline{D}_T) - Y(\overline{D}'_T)\right\}}{\frac{1}{T}\sum_{t=1}^{T}\left|\mathbb{E}\left(D_t - D'_t\right)\right|}}$$

- $\mathbb{E}\left\{Y(\overline{D}_T) - Y(\overline{D}'_T)\right\}$: same as before
- $T^{-1}\sum_{t=1}^{T}\left|\mathbb{E}\left(D_t - D'_t\right)\right|$: *average absolute per-timepoint change in the number of treatments*

There are multiple options for the denominator

1. $T^{-1}\sum_{t=1}^{T}\mathbb{P}\left(D_t \neq D'_t\right)$: *Average per-timepoint probability of switching treatment*
2. *Joint distributional distance* between dist. of $\overline{D}_T$ and dist. of $\overline{D}'_T$ (e.g., f-divergence, optimal transport metric).

Theorem: identification

f_t: replace $\mathbb{P}\{A_t(\overline{D}_{t-1}) = a_t \mid H_t(\overline{D}_{t-1})\}$ by $\mathbb{P}(A_t = a_t \mid H_t)$ in f_t^c

Suppose

- the **NPSEM and strong SR** hold and
- $\mathbb{P}(A_t = a_t \mid H_t) = 0 \implies f_t(a_t; H_t) = 0$.

Then,

$$\mathbb{E}\{Y(\overline{D}_T)\} = \overbrace{\sum_{\overline{b}_T \in \{0,1\}^T} \mathbb{E}\left\{\mathbb{E}(Y \mid \overline{A}_T = \overline{b}_T,\ \overline{X}_T) \prod_{t=1}^{T} Q_t(b_t \mid \overline{b}_{t-1}, \overline{X}_t)\right\}}^{\textbf{extended g-formula}}$$

$$= \left.\mathbb{E}\left[Y \prod_{t=1}^{T} \frac{Q_t(A_t \mid H_t)}{\mathbb{P}(A_t \mid H_t)}\right]\right\}\textbf{IPW}$$

where $\boxed{Q_t(a_t \mid h_t) = \mathbb{P}(A_t = a_t \mid h_t) + f_t(a_t; h_t)\{1 - \mathbb{P}(A_t = a_t \mid h_t)\}}$

$f_t(a_t; h_t) \approx 1 \implies Q_t(a_t \mid h_t) \approx 1$
$f_t(a_t; h_t) \approx 0 \implies Q_t(a_t \mid h_t) \approx \mathbb{P}(A_t = a_t \mid h_t)$

$\mathbb{E}(D_t)$ is ID'd in similar manner

Theorem: identification

f_t: replace $\mathbb{P}\{A_t(\overline{D}_{t-1}) = a_t \mid H_t(\overline{D}_{t-1})\}$ by $\mathbb{P}(A_t = a_t \mid H_t)$ in f_t^c

Suppose

- the **NPSEM and strong SR** hold and
- $\mathbb{P}(A_t = a_t \mid H_t) = 0 \implies f_t(a_t; H_t) = 0.$

Then,

$$\mathbb{E}\{Y(\overline{D}_T)\} = \overbrace{\sum_{\overline{b}_T \in \{0,1\}^T} \mathbb{E}\left\{\mathbb{E}(Y \mid \overline{A}_T = \overline{b}_T,\ \overline{X}_T) \prod_{t=1}^{T} Q_t(b_t \mid \overline{b}_{t-1}, \overline{X}_t)\right\}}^{\text{extended g-formula}}$$

$$= \left.\mathbb{E}\left[Y \prod_{t=1}^{T} \frac{Q_t(A_t \mid H_t)}{\mathbb{P}(A_t \mid H_t)}\right]\right\}\textbf{IPW}$$

where $\boxed{Q_t(a_t \mid h_t) = \mathbb{P}(A_t = a_t \mid h_t) + f_t(a_t; h_t)\{1 - \mathbb{P}(A_t = a_t \mid h_t)\}}$

$f_t(a_t; h_t) \approx 1 \implies Q_t(a_t \mid h_t) \approx 1$
$f_t(a_t; h_t) \approx 0 \implies Q_t(a_t \mid h_t) \approx \mathbb{P}(A_t = a_t \mid h_t)$

$\mathbb{E}(D_t)$ is ID'd in similar manner

Theorem: identification

f_t: replace $\mathbb{P}\{A_t(\overline{D}_{t-1}) = a_t \mid H_t(\overline{D}_{t-1})\}$ by $\mathbb{P}(A_t = a_t \mid H_t)$ in f_t^c

Suppose

- the **NPSEM and strong SR** hold and
- $\mathbb{P}(A_t = a_t \mid H_t) = 0 \implies f_t(a_t; H_t) = 0.$

Then,

$$\mathbb{E}\{Y(\overline{D}_T)\} = \overbrace{\sum_{\overline{b}_T \in \{0,1\}^T} \mathbb{E}\left\{\mathbb{E}(Y \mid \overline{A}_T = \overline{b}_T,\ \overline{X}_T) \prod_{t=1}^{T} Q_t(b_t \mid \overline{b}_{t-1}, \overline{X}_t)\right\}}^{\text{extended g-formula}}$$

$$= \underbrace{\mathbb{E}\left[Y \prod_{t=1}^{T} \frac{Q_t(A_t \mid H_t)}{\mathbb{P}(A_t \mid H_t)}\right]}_{\text{IPW}}$$

where $\boxed{Q_t(a_t \mid h_t) = \mathbb{P}(A_t = a_t \mid h_t) + f_t(a_t; h_t)\{1 - \mathbb{P}(A_t = a_t \mid h_t)\}}$

$f_t(a_t; h_t) \approx 1 \implies Q_t(a_t \mid h_t) \approx 1$
$f_t(a_t; h_t) \approx 0 \implies Q_t(a_t \mid h_t) \approx \mathbb{P}(A_t = a_t \mid h_t)$

$\mathbb{E}(D_t)$ is ID'd in similar manner

Theorem: identification

f_t: replace $\mathbb{P}\{A_t(\overline{D}_{t-1}) = a_t \mid H_t(\overline{D}_{t-1})\}$ by $\mathbb{P}(A_t = a_t \mid H_t)$ in f_t^c

Suppose

- the **NPSEM and strong SR** hold and
- $\mathbb{P}(A_t = a_t \mid H_t) = 0 \implies f_t(a_t; H_t) = 0$.

Then,

$$\mathbb{E}\{Y(\overline{D}_T)\} = \overbrace{\sum_{\overline{b}_T \in \{0,1\}^T} \mathbb{E}\left\{\mathbb{E}(Y \mid \overline{A}_T = \overline{b}_T,\ \overline{X}_T) \prod_{t=1}^{T} Q_t(b_t \mid \overline{b}_{t-1}, \overline{X}_t)\right\}}^{\text{extended g-formula}}$$

$$= \underbrace{\mathbb{E}\left[Y \prod_{t=1}^{T} \frac{Q_t(A_t \mid H_t)}{\mathbb{P}(A_t \mid H_t)}\right]}_{\textbf{IPW}}$$

where $\boxed{Q_t(a_t \mid h_t) = \mathbb{P}(A_t = a_t \mid h_t) + f_t(a_t; h_t)\{1 - \mathbb{P}(A_t = a_t \mid h_t)\}}$

$f_t(a_t; h_t) \approx 1 \implies Q_t(a_t \mid h_t) \approx 1$
$f_t(a_t; h_t) \approx 0 \implies Q_t(a_t \mid h_t) \approx \mathbb{P}(A_t = a_t \mid h_t)$

$\mathbb{E}(D_t)$ is ID'd in similar manner

Theorem: identification

f_t: replace $\mathbb{P}\{A_t(\overline{D}_{t-1}) = a_t \mid H_t(\overline{D}_{t-1})\}$ by $\mathbb{P}(A_t = a_t \mid H_t)$ in f_t^c

Suppose

- the **NPSEM and strong SR** hold and
- $\mathbb{P}(A_t = a_t \mid H_t) = 0 \implies f_t(a_t; H_t) = 0$.

Then,

$$\mathbb{E}\{Y(\overline{D}_T)\} = \overbrace{\sum_{\overline{b}_T \in \{0,1\}^T} \mathbb{E}\left\{\mathbb{E}(Y \mid \overline{A}_T = \overline{b}_T,\ \overline{X}_T) \prod_{t=1}^{T} Q_t(b_t \mid \overline{b}_{t-1}, \overline{X}_t)\right\}}^{\textbf{extended g-formula}}$$

$$= \underbrace{\mathbb{E}\left[Y \prod_{t=1}^{T} \frac{Q_t(A_t \mid H_t)}{\mathbb{P}(A_t \mid H_t)}\right]}_{\textbf{IPW}}$$

where $\boxed{Q_t(a_t \mid h_t) = \mathbb{P}(A_t = a_t \mid h_t) + f_t(a_t; h_t)\{1 - \mathbb{P}(A_t = a_t \mid h_t)\}}$

$f_t(a_t; h_t) \approx 1 \implies Q_t(a_t \mid h_t) \approx 1$
$f_t(a_t; h_t) \approx 0 \implies Q_t(a_t \mid h_t) \approx \mathbb{P}(A_t = a_t \mid h_t)$

$\mathbb{E}(D_t)$ is ID'd in similar manner

Theorem: identification

f_t: replace $\mathbb{P}\{A_t(\overline{D}_{t-1}) = a_t \mid H_t(\overline{D}_{t-1})\}$ by $\mathbb{P}(A_t = a_t \mid H_t)$ in f_t^c

Suppose

- the **NPSEM and strong SR** hold and
- $\mathbb{P}(A_t = a_t \mid H_t) = 0 \implies f_t(a_t; H_t) = 0.$

Then,

$$\mathbb{E}\{Y(\overline{D}_T)\} = \overbrace{\sum_{\overline{b}_T \in \{0,1\}^T} \mathbb{E}\left\{\mathbb{E}(Y \mid \overline{A}_T = \overline{b}_T,\ \overline{X}_T) \prod_{t=1}^{T} Q_t\left(b_t \mid \overline{b}_{t-1}, \overline{X}_t\right)\right\}}^{\textbf{extended g-formula}}$$

$$= \left.\mathbb{E}\left[Y \prod_{t=1}^{T} \frac{Q_t(A_t \mid H_t)}{\mathbb{P}(A_t \mid H_t)}\right]\right\}\textbf{IPW}$$

where $\boxed{Q_t(a_t \mid h_t) = \mathbb{P}(A_t = a_t \mid h_t) + f_t(a_t; h_t)\{1 - \mathbb{P}(A_t = a_t \mid h_t)\}}$

$f_t(a_t; h_t) \approx 1 \implies Q_t(a_t \mid h_t) \approx 1$
$f_t(a_t; h_t) \approx 0 \implies Q_t(a_t \mid h_t) \approx \mathbb{P}(A_t = a_t \mid h_t)$

$\mathbb{E}(D_t)$ is ID'd in similar manner

Theorem: identification

f_t: replace $\mathbb{P}\{A_t(\overline{D}_{t-1}) = a_t \mid H_t(\overline{D}_{t-1})\}$ by $\mathbb{P}(A_t = a_t \mid H_t)$ in f_t^c

Suppose

- the **NPSEM and strong SR** hold and
- $\mathbb{P}(A_t = a_t \mid H_t) = 0 \implies f_t(a_t; H_t) = 0.$

Then,

$$\mathbb{E}\{Y(\overline{D}_T)\} = \overbrace{\sum_{\overline{b}_T \in \{0,1\}^T} \mathbb{E}\left\{\mathbb{E}(Y \mid \overline{A}_T = \overline{b}_T,\ \overline{X}_T) \prod_{t=1}^{T} Q_t\left(b_t \mid \overline{b}_{t-1}, \overline{X}_t\right)\right\}}^{\text{extended g-formula}}$$

$$= \underbrace{\mathbb{E}\left[Y \prod_{t=1}^{T} \frac{Q_t(A_t \mid H_t)}{\mathbb{P}(A_t \mid H_t)}\right]}\text{IPW}$$

where $\boxed{Q_t(a_t \mid h_t) = \mathbb{P}(A_t = a_t \mid h_t) + f_t(a_t; h_t)\{1 - \mathbb{P}(A_t = a_t \mid h_t)\}}$

$f_t(a_t; h_t) \approx 1 \implies Q_t(a_t \mid h_t) \approx 1$
$f_t(a_t; h_t) \approx 0 \implies Q_t(a_t \mid h_t) \approx \mathbb{P}(A_t = a_t \mid h_t)$

$\mathbb{E}(D_t)$ is ID'd in similar manner

Estimation

(in two slides)

Estimating smooth flip effects

1. $\mathbb{E}\{Y(\overline{D}_T)\} = \mathbb{E}\left\{Y \prod_{t=1}^{T} \frac{Q_t(A_t|H_t)}{\mathbb{P}(A_t|H_t)}\right\}$
2. $Q_t(a_t \mid h_t) = \mathbb{P}(A_t = a_t \mid h_t) + f_t(a_t; h_t)\{1 - \mathbb{P}(A_t = a_t \mid h_t)\}$

Let $\boxed{f_t(a_t; h_t) = s_t\{\mathbb{P}(A_t = a_t \mid h_t)\}}$.

$s_t(\cdot)$ non-smooth (e.g., trimming: $\mathbb{1}(\varepsilon < \mathbb{P}(A_t = a_t \mid h_t) < 1 - \varepsilon)$)
$\implies \widehat{\mathbb{P}}(A_t = a_t \mid h_t)$ drives convergence

$s_t(\cdot)$ smooth (e.g., smooth trimming, overlap): flip effects are pathwise differentiable $\implies$ **efficient influence function** (EIF)-based estimators

For smooth $s_t(\cdot)$, we derive the new EIF (plug-in plus weighted residuals)

Inspires two one-step estimators (for $\mathbb{E}\{Y(\overline{D}_T)\}$ and $\mathbb{E}(D_t)$):

1. **Multiply** robust-style
2. **Sequentially doubly** robust-style (debias pseudo-outcome in multiply robust-style estimator)

Estimating smooth flip effects

1. $\mathbb{E}\{Y(\overline{D}_T)\} = \mathbb{E}\left\{Y \prod_{t=1}^{T} \frac{Q_t(A_t|H_t)}{\mathbb{P}(A_t|H_t)}\right\}$
2. $Q_t(a_t \mid h_t) = \mathbb{P}(A_t = a_t \mid h_t) + f_t(a_t; h_t)\{1 - \mathbb{P}(A_t = a_t \mid h_t)\}$

Let $\boxed{f_t(a_t; h_t) = s_t\{\mathbb{P}(A_t = a_t \mid h_t)\}}$.

$s_t(\cdot)$ non-smooth (e.g., trimming: $\mathbb{1}(\varepsilon < \mathbb{P}(A_t = a_t \mid h_t) < 1 - \varepsilon)$)
$\implies \widehat{\mathbb{P}}(A_t = a_t \mid h_t)$ drives convergence

$s_t(\cdot)$ smooth (e.g., smooth trimming, overlap): flip effects are pathwise differentiable $\implies$ **efficient influence function** (EIF)-based estimators

For smooth $s_t(\cdot)$, we derive the new EIF (plug-in plus weighted residuals)

Inspires two one-step estimators (for $\mathbb{E}\{Y(\overline{D}_T)\}$ and $\mathbb{E}(D_t)$):

1. **Multiply** robust-style
2. **Sequentially doubly** robust-style (debias pseudo-outcome in multiply robust-style estimator)

Estimating smooth flip effects

1. $\mathbb{E}\{Y(\overline{D}_T)\} = \mathbb{E}\left\{ Y \prod_{t=1}^{T} \frac{Q_t(A_t|H_t)}{\mathbb{P}(A_t|H_t)} \right\}$
2. $Q_t(a_t \mid h_t) = \mathbb{P}(A_t = a_t \mid h_t) + f_t(a_t; h_t)\{1 - \mathbb{P}(A_t = a_t \mid h_t)\}$

Let $\boxed{f_t(a_t; h_t) = s_t\{\mathbb{P}(A_t = a_t \mid h_t)\}}$.

$s_t(\cdot)$ non-smooth (e.g., trimming: $\mathbb{1}(\varepsilon < \mathbb{P}(A_t = a_t \mid h_t) < 1 - \varepsilon)$)
$\implies \widehat{\mathbb{P}}(A_t = a_t \mid h_t)$ drives convergence

$s_t(\cdot)$ smooth (e.g., smooth trimming, overlap): flip effects are pathwise differentiable $\implies$ **efficient influence function** (EIF)-based estimators

For smooth $s_t(\cdot)$, we derive the new EIF (plug-in plus weighted residuals)

Inspires two one-step estimators (for $\mathbb{E}\{Y(\overline{D}_T)\}$ and $\mathbb{E}(D_t)$):

1. **Multiply** robust-style
2. **Sequentially doubly** robust-style (debias pseudo-outcome in multiply robust-style estimator)

Estimating smooth flip effects

1. $\mathbb{E}\{Y(\overline{D}_T)\} = \mathbb{E}\left\{Y \prod_{t=1}^{T} \frac{Q_t(A_t|H_t)}{\mathbb{P}(A_t|H_t)}\right\}$
2. $Q_t(a_t \mid h_t) = \mathbb{P}(A_t = a_t \mid h_t) + f_t(a_t; h_t)\{1 - \mathbb{P}(A_t = a_t \mid h_t)\}$

Let $\boxed{f_t(a_t; h_t) = s_t\{\mathbb{P}(A_t = a_t \mid h_t)\}}$.

$s_t(\cdot)$ non-smooth (e.g., trimming: $\mathbb{1}(\varepsilon < \mathbb{P}(A_t = a_t \mid h_t) < 1 - \varepsilon)$)
$\implies \widehat{\mathbb{P}}(A_t = a_t \mid h_t)$ drives convergence

$s_t(\cdot)$ smooth (e.g., smooth trimming, overlap): flip effects are pathwise differentiable $\implies$ **efficient influence function** (EIF)-based estimators

For smooth $s_t(\cdot)$, we derive the new EIF (plug-in plus weighted residuals)

Inspires two one-step estimators (for $\mathbb{E}\{Y(\overline{D}_T)\}$ and $\mathbb{E}(D_t)$):

1. **Multiply** robust-style
2. **Sequentially doubly** robust-style (debias pseudo-outcome in multiply robust-style estimator)

Estimating smooth flip effects

1. $\mathbb{E}\{Y(\overline{D}_T)\} = \mathbb{E}\left\{ Y \prod_{t=1}^{T} \frac{Q_t(A_t|H_t)}{\mathbb{P}(A_t|H_t)} \right\}$
2. $Q_t(a_t \mid h_t) = \mathbb{P}(A_t = a_t \mid h_t) + f_t(a_t; h_t)\{1 - \mathbb{P}(A_t = a_t \mid h_t)\}$

Let $\boxed{f_t(a_t; h_t) = s_t\{\mathbb{P}(A_t = a_t \mid h_t)\}}$.

$s_t(\cdot)$ non-smooth (e.g., trimming: $\mathbb{1}(\varepsilon < \mathbb{P}(A_t = a_t \mid h_t) < 1 - \varepsilon)$)
$\implies \widehat{\mathbb{P}}(A_t = a_t \mid h_t)$ drives convergence

$s_t(\cdot)$ smooth (e.g., smooth trimming, overlap): flip effects are pathwise differentiable $\implies$ **efficient influence function** (EIF)-based estimators

For smooth $s_t(\cdot)$, we derive the new EIF (plug-in plus weighted residuals)

Inspires two one-step estimators (for $\mathbb{E}\{Y(\overline{D}_T)\}$ and $\mathbb{E}(D_t)$):

1. **Multiply** robust-style
2. **Sequentially doubly** robust-style (debias pseudo-outcome in multiply robust-style estimator)

Estimating smooth flip effects

1. $\mathbb{E}\{Y(\overline{D}_T)\} = \mathbb{E}\left\{Y \prod_{t=1}^{T} \frac{Q_t(A_t|H_t)}{\mathbb{P}(A_t|H_t)}\right\}$
2. $Q_t(a_t \mid h_t) = \mathbb{P}(A_t = a_t \mid h_t) + f_t(a_t; h_t)\{1 - \mathbb{P}(A_t = a_t \mid h_t)\}$

Let $\boxed{f_t(a_t; h_t) = s_t\{\mathbb{P}(A_t = a_t \mid h_t)\}}$.

$s_t(\cdot)$ non-smooth (e.g., trimming: $\mathbb{1}(\varepsilon < \mathbb{P}(A_t = a_t \mid h_t) < 1 - \varepsilon)$)
$\implies \widehat{\mathbb{P}}(A_t = a_t \mid h_t)$ drives convergence

$s_t(\cdot)$ smooth (e.g., smooth trimming, overlap): flip effects are pathwise differentiable $\implies$ **efficient influence function** (EIF)-based estimators

For smooth $s_t(\cdot)$, we derive the new EIF (plug-in plus weighted residuals)

Inspires two one-step estimators (for $\mathbb{E}\{Y(\overline{D}_T)\}$ and $\mathbb{E}(D_t)$):

1. **Multiply** robust-style
2. **Sequentially doubly** robust-style (debias pseudo-outcome in multiply robust-style estimator)

Estimating smooth flip effects

1. $\mathbb{E}\{Y(\overline{D}_T)\} = \mathbb{E}\left\{ Y \prod_{t=1}^{T} \frac{Q_t(A_t|H_t)}{\mathbb{P}(A_t|H_t)} \right\}$
2. $Q_t(a_t \mid h_t) = \mathbb{P}(A_t = a_t \mid h_t) + f_t(a_t; h_t)\{1 - \mathbb{P}(A_t = a_t \mid h_t)\}$

Let $\boxed{f_t(a_t; h_t) = s_t\{\mathbb{P}(A_t = a_t \mid h_t)\}}$.

$s_t(\cdot)$ non-smooth (e.g., trimming: $\mathbb{1}(\varepsilon < \mathbb{P}(A_t = a_t \mid h_t) < 1 - \varepsilon)$)
$\implies \widehat{\mathbb{P}}(A_t = a_t \mid h_t)$ drives convergence

$s_t(\cdot)$ smooth (e.g., smooth trimming, overlap): flip effects are pathwise differentiable $\implies$ **efficient influence function** (EIF)-based estimators

For smooth $s_t(\cdot)$, we derive the new EIF (plug-in plus weighted residuals)

Inspires two one-step estimators (for $\mathbb{E}\{Y(\overline{D}_T)\}$ and $\mathbb{E}(D_t)$):

1. **Multiply** robust-style
2. **Sequentially doubly** robust-style (debias pseudo-outcome in multiply robust-style estimator)

Estimating smooth flip effects

1. $\mathbb{E}\{Y(\overline{D}_T)\} = \mathbb{E}\left\{Y \prod_{t=1}^{T} \frac{Q_t(A_t|H_t)}{\mathbb{P}(A_t|H_t)}\right\}$
2. $Q_t(a_t \mid h_t) = \mathbb{P}(A_t = a_t \mid h_t) + f_t(a_t; h_t)\{1 - \mathbb{P}(A_t = a_t \mid h_t)\}$

Let $\boxed{f_t(a_t; h_t) = s_t\{\mathbb{P}(A_t = a_t \mid h_t)\}}$.

$s_t(\cdot)$ non-smooth (e.g., trimming: $\mathbb{1}(\varepsilon < \mathbb{P}(A_t = a_t \mid h_t) < 1 - \varepsilon)$)
$\implies \widehat{\mathbb{P}}(A_t = a_t \mid h_t)$ drives convergence

$s_t(\cdot)$ smooth (e.g., smooth trimming, overlap): flip effects are pathwise differentiable $\implies$ **efficient influence function** (EIF)-based estimators

For smooth $s_t(\cdot)$, we derive the new EIF (plug-in plus weighted residuals)

Inspires two one-step estimators (for $\mathbb{E}\{Y(\overline{D}_T)\}$ and $\mathbb{E}(D_t)$):

1. **Multiply** robust-style
2. **Sequentially doubly** robust-style (debias pseudo-outcome in multiply robust-style estimator)

Bias bounds

1. **Multiply robust-style estimator** has new result:
   Bias is **minimum** of two errors
   → Díaz et al. [2023] & Kennedy [2019]: unroll into future and past
   → $2(T+1)$ multiply robust-style bound
2. **Sequentially DR-style** estimator is **first-of-its-kind** where Q_t depends on unknown propensity score

Bias bounds $\implies$ weak convergence under nonparametric (ML) conditions:

Theorem, informal: Let ψ denote flip effect, π_t prop. score at t, $\widetilde{m}_t$ denote seq. reg. at t with estimated pseudo-outcome. Then,

$$\boxed{\left|\mathbb{E}\left(\widehat{\psi}_{sdr} - \psi\right)\right| \lesssim \sum_{t=1}^{T} \|\widehat{\pi}_t - \pi_t\| \left(\|\widehat{m}_t - \widetilde{m}_t\| + \|\widehat{\pi}_t - \pi_t\| \right).}$$

$\overbrace{\|\widehat{\pi}_t - \pi_t\| = o_{\mathbb{P}}(n^{-1/4}) = \|\widehat{m}_t - \widetilde{m}_t\|}^{\text{ML conv. rates}} \implies \sqrt{n}$−consistency & asymptotic normality

Bias bounds

1. **Multiply robust-style estimator** has new result:
   Bias is **minimum** of two errors
   $\rightarrow$ Díaz et al. [2023] & Kennedy [2019]: unroll into future and past
   $\rightarrow$ $2(T+1)$ multiply robust-style bound
2. **Sequentially DR-style** estimator is **first-of-its-kind** where Q_t depends on unknown propensity score

Bias bounds $\implies$ weak convergence under nonparametric (ML) conditions:

Theorem, informal: Let ψ denote flip effect, π_t prop. score at t, $\widetilde{m}_t$ denote seq. reg. at t with estimated pseudo-outcome. Then,

$$\boxed{\left|\mathbb{E}\left(\widehat{\psi}_{sdr} - \psi\right)\right| \lesssim \sum_{t=1}^{T} \|\widehat{\pi}_t - \pi_t\| \Big(\|\widehat{m}_t - \widetilde{m}_t\| + \|\widehat{\pi}_t - \pi_t\| \Big).}$$

$\overbrace{\|\widehat{\pi}_t - \pi_t\| = o_{\mathbb{P}}(n^{-1/4}) = \|\widehat{m}_t - \widetilde{m}_t\|}^{\text{ML conv. rates}} \implies \sqrt{n}$–consistency & asymptotic normality

Bias bounds

1. **Multiply robust-style estimator** has new result:
   Bias is **minimum** of two errors
   → Díaz et al. [2023] & Kennedy [2019]: unroll into future and past
   → $2(T+1)$ multiply robust-style bound
2. **Sequentially DR-style** estimator is **first-of-its-kind** where Q_t depends on unknown propensity score

Bias bounds $\implies$ weak convergence under nonparametric (ML) conditions:

Theorem, informal: Let ψ denote flip effect, π_t prop. score at t, $\widetilde{m}_t$ denote seq. reg. at t with estimated pseudo-outcome. Then,

$$\boxed{\left|\mathbb{E}\left(\widehat{\psi}_{sdr} - \psi\right)\right| \lesssim \sum_{t=1}^{T} \|\widehat{\pi}_t - \pi_t\| \Big(\|\widehat{m}_t - \widetilde{m}_t\| + \|\widehat{\pi}_t - \pi_t\|\Big).}$$

$\overbrace{\|\widehat{\pi}_t - \pi_t\| = o_{\mathbb{P}}(n^{-1/4}) = \|\widehat{m}_t - \widetilde{m}_t\|}^{\text{ML conv. rates}} \implies \sqrt{n}$−consistency & asymptotic normality

Bias bounds

1. **Multiply robust-style estimator** has new result:
   Bias is **minimum** of two errors
   $\rightarrow$ Díaz et al. [2023] & Kennedy [2019]: unroll into future and past
   $\rightarrow 2(T+1)$ multiply robust-style bound
2. **Sequentially DR-style** estimator is **first-of-its-kind** where Q_t depends on unknown propensity score

Bias bounds $\implies$ weak convergence under nonparametric (ML) conditions:

Theorem, informal: Let ψ denote flip effect, π_t prop. score at t, $\widetilde{m}_t$ denote seq. reg. at t with estimated pseudo-outcome. Then,

$$\boxed{\left|\mathbb{E}\left(\widehat{\psi}_{sdr} - \psi\right)\right| \lesssim \sum_{t=1}^{T} \|\widehat{\pi}_t - \pi_t\| \Big(\|\widehat{m}_t - \widetilde{m}_t\| + \|\widehat{\pi}_t - \pi_t\|\Big).}$$

$\overbrace{\|\widehat{\pi}_t - \pi_t\| = o_{\mathbb{P}}(n^{-1/4}) = \|\widehat{m}_t - \widetilde{m}_t\|}^{\text{ML conv. rates}} \implies \sqrt{n}$–consistency & asymptotic normality

Bias bounds

1. **Multiply robust-style estimator** has new result: **Bias** is **minimum** of two errors
   $\rightarrow$ Díaz et al. [2023] & Kennedy [2019]: unroll into future and past
   $\rightarrow 2(T+1)$ multiply robust-style bound
2. **Sequentially DR-style** estimator is **first-of-its-kind** where Q_t depends on unknown propensity score

Bias bounds $\implies$ weak convergence under nonparametric (ML) conditions:

Theorem, informal: Let ψ denote flip effect, π_t prop. score at t, $\widetilde{m}_t$ denote seq. reg. at t with estimated pseudo-outcome. Then,

$$\boxed{\left|\mathbb{E}\left(\widehat{\psi}_{sdr}-\psi\right)\right| \lesssim \sum_{t=1}^{T} \|\widehat{\pi}_t-\pi_t\|\Big(\|\widehat{m}_t-\widetilde{m}_t\|+\|\widehat{\pi}_t-\pi_t\|\Big).}$$

$\overbrace{\|\widehat{\pi}_t-\pi_t\| = o_{\mathbb{P}}(n^{-1/4}) = \|\widehat{m}_t-\widetilde{m}_t\|}^{\text{ML conv. rates}} \implies \sqrt{n}-$consistency & asymptotic normality

Thank you!

Recap:

- $T = 1$:
  WATEs $\equiv$ interventional flip effects
- $T > 1$: more complicated
  Flip ints allow weighting/trimming on non-baseline covariates; robust to positivity violations
  Can identify and estimate longitudinal interventional flip effects
- Efficient estimation of longitudinal interventional flip effects:
  (1) multiply robust and
  (2) sequentially doubly robust

hadera01@nyu.edu

Prelim draft:

alecmcclean.github.io/files/long-weight-short.pdf

Iván Díaz, Nicholas Williams, Katherine L Hoffman, and Edward J Schenck. Nonparametric causal effects based on longitudinal modified treatment policies. *Journal of the American Statistical Association*, 118(542):846–857, 2023.

Sebastian Haneuse and Andrea Rotnitzky. Estimation of the effect of interventions that modify the received treatment. *Statistics in medicine*, 32(30):5260–5277, 2013.

Edward H Kennedy. Nonparametric causal effects based on incremental propensity score interventions. *Journal of the American Statistical Association*, 114(526):645–656, 2019.

Alexander W Levis, Edward H Kennedy, Alec McClean, Sivaraman Balakrishnan, and Larry Wasserman. Stochastic interventions, sensitivity analysis, and optimal transport. *arXiv preprint arXiv:2411.14285*, 2024.

Maya L Petersen, Kristin E Porter, Susan Gruber, Yue Wang, and Mark J Van Der Laan. Diagnosing and responding to violations in the positivity assumption. *Statistical methods in medical research*, 21(1):31–54, 2012.

Xiang Zhou and Aleksei Opacic. Marginal interventional effects. *arXiv preprint arXiv:2206.10717*, 2022.

Additional remarks

1. Robustness to positivity violations: **Smooth flip effects are alternative to IPSIs that target a specific regime**
2. **Relax dependence on natural value of treatment:**

$$D_t = \mathbb{1}\Big[V_t \leq \mathbb{1}(a_t = 1)f_t^c + \big(1 - f_t^c\big)\mathbb{P}\{A_t(\overline{D}_{t-1}) = 1 \mid H_t(\overline{D}_{t-1})\}\Big]$$

   - Can avoid practical issues — we may not observe natural value of treatment
   - Identification only requires standard SR
3. Flip interventions inspired by **maximally coupled policies** [Levis et al., 2024]
4. **Single-timepoint WATEs:**
   - Conditional on $\widehat{f}$, any WATE can be reinterpreted via flip ints post-hoc
   - Our analysis $\implies$ DR-style estimation for all single-timepoint WATEs w/ smooth weights.

Drawbacks with direct weighting; or, why we need flip interventions

Longitudinal interventional flip effects do not yield, e.g.,

$$\mathbb{E}\left(\frac{\{Y(\overline{a}_T) - Y(\overline{a}'_T)\}\prod_{t=1}^{T} f_t^c\{a_t; H_t(\overline{a}_{t-1})\} f_t^c\{a'_t; H_t(\overline{a}'_{t-1})\}}{\mathbb{E}\left[\prod_{t=1}^{T} f_t^c\{a_t; H_t(\overline{a}_{t-1})\} f_t^c\{a'_t; H_t(\overline{a}'_{t-1})\}\right]}\right) \quad (1)$$

Weighted average treatment effect of $\overline{a}_T$ versus $\overline{a}'_T$, where weights based on counterfactual propensity scores under each regime.

Why? because $\overline{D}_t$ affects $X_{t+1}(\overline{D}_t), A_{t+1}(\overline{D}_t)$

However, (1) is a bad target!

Interventions that yield (1) are:

1. Cross-world: at $t = 2$, weight using propensity scores under $D_1 = 1$ and $D_1 = 0$; non-falsifiable
2. Future-dependent: at $t = 1$, weight using future propensity scores at $t = 2$

(this justifies the caution typically advised)