Day 4, Lecture 1

Identification of effects of time-dependent treatment interventions

Identifying effects of time-dependent treatment interventions

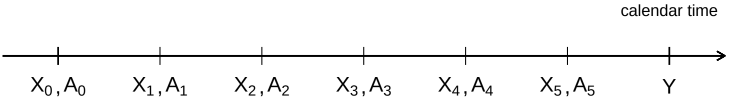
In this lecture, our goal is to:

1. Identify and list the key identification assumptions necessary for a causal interpretation of parameters defined under dynamic treatment interventions, highlighting particularly on the challenges imposed by time-dependent confounding.
2. Explain the identification formulas in presence of time-dependent treatments and confounding, with a specific focus on the identification achieved through sequential regression steps.

Identifying effects of time-dependent treatment interventions

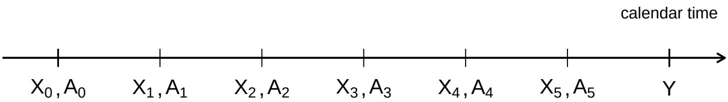
Longitudinal data structure:

- $O = (X_0, A_0, X_1, A_1, \ldots, X_K, A_K, Y) \in (\mathbb{R}^d \times \{0,1\})^K \times \{0,1\}$
- Covariates $X = (X_0, X_1, \ldots, X_K)$ change over time
- Treatment decisions $A = (A_0, A_1, \ldots, A_K)$ are updated over time
- Covariates and treatment decisions interact in complex ways



NB: For now keeping right-censoring (and competing risks) out of the picture.

Identifying effects of time-dependent treatment interventions



Counterfactual outcomes

$$Y^{A_{\mathbf{0}}=a^*_{\mathbf{0}}, A_{\mathbf{1}}=a^*_{\mathbf{1}}, \dots, A_K=a^*_K}, \qquad \text{for,} \quad a^*_0, \dots, a^*_K \in \{0,1\}$$

= defined by a sequence of treatment decision rules that we choose.

also called:

- hypothetical treatment interventions
- hypothetical treatment strategies
- hypothetical treatment regimes

NB: For now keeping right-censoring (and competing risks) out of the picture.

Overview

1. Identifying assumptions
   - No unmeasured confounding and positivity
2. Identification formulas
   - Inverse probability weighting
   - Sequential regression (iterated expectations)
3. Practical 1

Identifying assumptions

Identifying effects of time-dependent treatment interventions

Identification of $\mathbb{E}[Y^{A_0=a_0^*, A_1=a_1^*, \ldots, A_K=a_K^*}]$.

1. Consistency: $Y^{A_0=a_0^*, A_1=a_1^*, \ldots, A_K=a_K^*} = Y$

   if $A_k = a_k^*$ for $k = 0, 1, \ldots, K$

2. Exchangeability: $Y^{A_0=a_0^*, A_1=a_1^*, \ldots, A_K=a_K^*} \perp\!\!\!\perp A_k \mid \bar{X}_k, \bar{A}_{k-1}$

   for $k = 0, 1, \ldots, K$

3. Positivity:
$$\prod_{k=0}^{K} \frac{1\{A_k = a_k^*\}}{P(A_k = a_k^* \mid \bar{X}_k, \bar{A}_{k-1})} < \infty$$

   for $k = 0, 1, \ldots, K$

Notation for histories of variables: $\bar{X}_k = (X_0, X_1, \ldots, X_k)$, $\bar{A}_k = (A_0, A_1, \ldots, A_k)$.

Identifying effects of time-dependent treatment interventions

Imposing a static regime, like 'always treat',

$$A_0 = 1, A_1 = 1, \ldots, A_K = 1$$

may not always be realistic (or even feasible).

Identifying effects of time-dependent treatment interventions

Imposing a static regime, like 'always treat',

$$A_0 = 1, A_1 = 1, \ldots, A_K = 1$$

may not always be realistic (or even feasible).

Example (Robins 1986) Effects of exposure of chemicals on employees: Static regimes cannot be identified since subjects can only be exposed when at work.

Identifying effects of time-dependent treatment interventions

Imposing a static regime, like 'always treat',

$$A_0 = 1, A_1 = 1, \ldots, A_K = 1$$

may not always be realistic (or even feasible).

Example (Robins 1986) Effects of exposure of chemicals on employees: Static regimes cannot be identified since subjects can only be exposed when at work.

Another example Development of adverse effects or contraindications (e.g., pregnancy) can force a subject to stop an assigned treatment.

Identifying effects of time-dependent treatment interventions

But the positivity assumption dictates that the treatment level imposed by the intervention cannot in the observed data be deterministically assigned at any time point based on a subject's observed past.

3. Positivity:

$$\prod_{k=0}^{K} \frac{1\{A_k = a_k^*\}}{P(A_k = a_k^* \mid \bar{X}_k, \bar{A}_{k-1})} < \infty$$

for $k = 0, 1, \ldots, K$

Identifying effects of time-dependent treatment interventions

What we can do $\Rightarrow$ change the question/intervention.

- 'Expose when at work'
- 'Treat until adverse event or contraindication happen'
- 'Initiate antidiabetic treatment when HbA1c level increases beyond some level'

Identifying effects of time-dependent treatment interventions

Dynamic treatment regimes

- A prespecified set of rules which assign treatment over time by responding to a patient's time-varying conditions.

Identifying effects of time-dependent treatment interventions

Dynamic treatment regimes

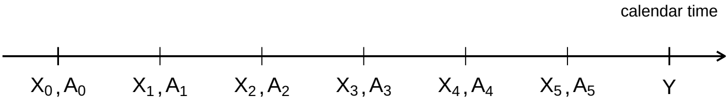
- A prespecified set of rules which assign treatment over time by responding to a patient's time-varying conditions.
- Mathematically, defined as function $\mathcal{S}_k(\bar{X}_k, \bar{A}_{k-1})$ that maps (a subset of) previous covariate/treatment values $\bar{X}_k, \bar{A}_{k-1}$ to a binary treatment assignment, e.g.,

$$\mathcal{S}_k(\bar{X}_k, \bar{A}_{k-1}) = \begin{cases} 1 & \text{if } X_k > \theta, \\ 0 & \text{if } X_k \leq \theta. \end{cases}$$

Identifying effects of time-dependent treatment interventions

2. Exchangeability: $Y^{A_0=a_0^*, A_1=a_1^*, \ldots, A_K=a_K^*} \perp\!\!\!\perp A_k \mid \bar{X}_k, \bar{A}_{k-1}$

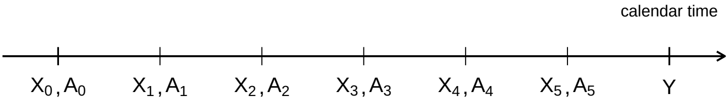
   for $k = 0, 1, \ldots, K$



- X_k may be affected by earlier treatment decisions $A_{k-1}, \ldots, A_1, A_0$.
- X_k may be a confounder for the effect of $A_k, A_{k+1}, \ldots, A_K$ on Y.

Identifying effects of time-dependent treatment interventions

2. Exchangeability: $Y^{A_0=a_0^*, A_1=a_1^*, \ldots, A_K=a_K^*} \perp\!\!\!\perp A_k \mid \bar{X}_k, \bar{A}_{k-1}$
   for $k = 0, 1, \ldots, K$



- X_k may be affected by earlier treatment decisions $A_{k-1}, \ldots, A_1, A_0$.
- X_k may be a confounder for the effect of $A_k, A_{k+1}, \ldots, A_K$ on Y.

time-dependent confounding

Identifying effects of time-dependent treatment interventions

In presence of time-dependent confounding, "standard methods" may cause bias

- Multiple regression
- Random effects models
- Time-dependent Cox regression

The problem is that:

- If we control for X_k in our model, we will not capture the effect from earlier treatment decisions $A_{k-1}, \ldots, A_1, A_0$ through X_k.
- But we have to control for X_k to assess the effect of $A_k, A_{k+1}, \ldots, A_K$ on Y.

Identifying effects of time-dependent treatment interventions

In Day 3, Practical 2, we considered the effects of different types of interventions:

1. The intention-to-treat (ITT) effect which only sets treatment at baseline and contrasts the two scenarios of being treated at baseline ($A_0 = 1$) and not being treated at baseline ($A_0 = 0$).
2. A static effect of being 'always treated' ($A_0 = A_1 = 1$) and 'never treated' ($A_0 = A_1 = 0$).
3. A dynamic effect of being treated at baseline ($A_0 = 1$) and only treated at follow-up if the adverse event has not happened, i.e., $X_{1,1} = 0$ — contrasted to being 'never treated' ($A_0 = A_1 = 0$).

Identifying effects of time-dependent treatment interventions

The true ITT average treatment effect:

```
ITT: -0.93%
```

The true static average treatment effect:

```
static: -6.33%
```

The true dynamic average treatment effect:

```
dynamic: -5.07%
```

Identifying effects of time-dependent treatment interventions

We considered two 'naive approaches' to estimation of the static effect:

Identifying effects of time-dependent treatment interventions

We considered two 'naive approaches' to estimation of the static effect:

1. Logistic regression of the outcome regressed on all treatment variables and covariates: Contrast means of the predictions under $A_0 = A_1 = 1$ to the mean of the predictions under $A_0 = A_1 = 0$.

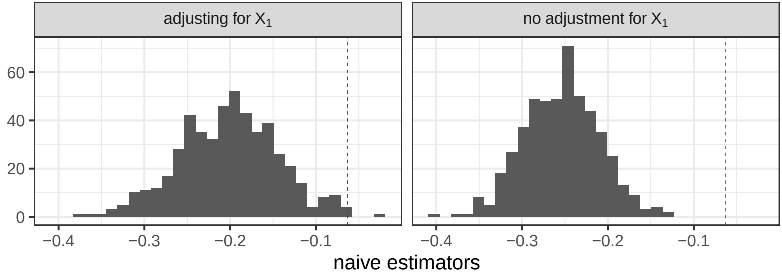
Identifying effects of time-dependent treatment interventions

We considered two 'naive approaches' to estimation of the static effect:

1. Logistic regression of the outcome regressed on all treatment variables and covariates: Contrast means of the predictions under $A_0 = A_1 = 1$ to the mean of the predictions under $A_0 = A_1 = 0$.
2. Logistic regression of the outcome regressed on baseline covariates and both treatment variables (leaving out follow-up covariates): Contrast means of the predictions under $A_0 = A_1 = 1$ to the mean of the predictions under $A_0 = A_1 = 0$.

Identifying effects of time-dependent treatment interventions

In a simulation study with $M = 500$ repetitions:



Both naive approaches give biased results — due to time-dependent confounding.

Identification formulas

Identification formulas

Factorization of the density p of $P \in \mathcal{M}$:[1]

$$p(o) = \mu_{X_0}(x_0)\pi_{A_0}(a \mid x_0) \prod_{k=1}^{K} \mu_{X_k}(x_k \mid \bar{x}_{k-1}, \bar{a}_{k-1})\pi_{A_k}(a_k \mid \bar{x}_k, \bar{a}_{k-1}) \times \mu_Y(y \mid \bar{x}_K, \bar{a}_K)$$

- μ_{X_0} is the marginal density of baseline covariates.
- π_{A_0} is the density of treatment at baseline.
- $\mu_{X_k}(x_k \mid \bar{x}_{k-1}, \bar{a}_{k-1})$ is the conditional density of X_k given the histories $\bar{X}_{k-1} = \bar{x}_{k-1}, \bar{A}_{k-1} = \bar{a}_{k-1},\ k = 1, \ldots, K$.
- $\pi_{A_k}(a_k \mid \bar{x}_k, \bar{a}_{k-1})$ is the conditional density of A_k given the histories $\bar{X}_k = \bar{x}_k, \bar{A}_{k-1} = \bar{a}_{k-1},\ k = 1, \ldots, K$.
- $\mu_Y(y \mid \bar{x}_K, \bar{a}_K)$ is the conditional density of Y given the histories $\bar{X}_K = \bar{x}_K, \bar{A}_K = \bar{a}_K$.

[1] Statistical model $\mathcal{M}$ for P_0 contains possible distributions P for the observed data O.

Identification formulas

Factorization of density allows us to write the expectation under P in terms of iterated integrals (Fubini's theorem):

$$
\begin{aligned}
\mathbb{E}_P[Y] &= \int_{\mathcal{O}} y p(o) d\nu(o) \\
&= \int_{\mathbb{R}^d} \sum_{a_0=0,1} \cdots \int_{\mathbb{R}^d} \sum_{a_K=0,1} \sum_{y=0,1} y \mu_Y(y \mid \bar{x}_K, \bar{a}_K) \\
&\qquad \pi_K(a_K \mid \bar{x}_K, \bar{a}_K) \mu_{X_K}(x_K \mid \bar{x}_{K-1}, \bar{a}_{K-1}) d\nu_{X_K}(x_K) \\
&\qquad\qquad \cdots \pi_0(a_0 \mid x_0) \mu_{X_0}(x_0) d\nu_{X_0}(x_0),
\end{aligned}
$$

for $P \in \mathcal{M}$.

Identification formulas

We want to identify the treatment-specific mean outcome:

$$\mathbb{E}_P\big[\,Y^{A_0=a_0^*,A_1=a_1^*,\dots,A_K=a_K^*}\,\big]$$

in terms of the observed data distribution

using the assumptions:

$$Y^{A_0=a_0^*,A_1=a_1^*,\dots,A_K=a_K^*} = Y \qquad \text{if } A_k = a_k^* \text{ for } k = 0, 1, \dots, K$$

$$Y^{A_0=a_0^*,A_1=a_1^*,\dots,A_K=a_K^*} \perp\!\!\!\perp A_k \mid \bar{X}_k, \bar{A}_{k-1}, \qquad \text{for } k = 0, 1, \dots, K$$

$$\prod_{k=0}^{K} \frac{1\{A_k = a_k^*\}}{P(A_k = a_k^* \mid \bar{X}_k, \bar{A}_{k-1})} < \infty, \qquad \text{for } k = 0, 1, \dots, K$$

Identification: g-formula

The claim is that:

$$
\begin{aligned}
&\mathbb{E}_P\big[Y^{A_0=a_0^*,A_1=a_1^*,\ldots,A_K=a_K^*}\big] \\
&\quad = \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} \sum_{y=0,1} y\mu_Y(y \mid \bar{x}_K, \bar{a}_K^*) \\
&\qquad\qquad \mu_{X_K}(x_K \mid \bar{x}_{K-1}, \bar{a}_{K-1}^*) d\nu_{X_K}(x_K) \cdots \mu_{X_0}(x_0) d\nu_{X_0}(x_0)
\end{aligned}
$$

Identification: g-formula

To show the claim from the previous slide, start from the right hand side:

1. By consistency, replace Y by $Y^{A_0=a_0^*,A_1=a_1^*,\dots,A_K=a_K^*}$ in the innermost integral:

$$\sum_{y=0,1} y\mu_Y(y \mid \bar{x}_K, \bar{a}_K^*) = \mathbb{E}_P[Y \mid \bar{X}_K = \bar{x}_K, \bar{A}_K = \bar{a}_K^*]$$

$$= \mathbb{E}_P[Y^{A_0=a_0^*,A_1=a_1^*,\dots,A_K=a_K^*} \mid \bar{X}_K = \bar{x}_K, \bar{A}_K = \bar{a}_K^*] \qquad (1)$$

2. Drop the last conditioning variable $A_K = a_K^*$ from the conditioning set by exchangeability, and then integrate out over L_K:

$$\int_{\mathbb{R}^d} \sum_{y=0,1} y\mu_Y(y \mid \bar{x}_K, \bar{a}_K^*)\mu_{X_K}(x_K \mid \bar{x}_{K-1}, \bar{a}_{K-1}^*)d\nu_{X_K}(x_K)$$

$$= \int_{\mathbb{R}^d} \mathbb{E}_P[Y^{A_0=a_0^*,A_1=a_1^*,\dots,A_K=a_K^*} \mid \bar{X}_K = \bar{x}_K, \bar{A}_{K-1}^* = \bar{a}_{K-1}^*]\mu_{X_K}(x_K \mid \bar{x}_{K-1}, \bar{a}_{K-1}^*)d\nu_{X_K}(x_K)$$

$$= \mathbb{E}_P[Y^{A_0=a_0^*,A_1=a_1^*,\dots,A_K=a_K^*} \mid \bar{X}_{K-1} = \bar{x}_{K-1}, \bar{A}_{K-1}^* = \bar{a}_{K-1}^*] \qquad (2)$$

3. Note that (2) is the same expression as (1), with K replaced by $K-1$.
4. Repeat 2. another $K-1$ times which in the end gives the left hand side from the previous slide.

Identification: IP-weighting

We have that:

$$\mathbb{E}_P[\,Y^{A_0=a_0^*,A_1=a_1^*,\ldots,A_K=a_K^*}\,]$$

$$= \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} \sum_{y=0,1} y \mu_Y(y \mid \bar{x}_K, \bar{a}_K^*)$$

$$\mu_{X_K}(x_K \mid \bar{x}_{K-1}, \bar{a}_{K-1}^*) d\nu_{X_K}(x_K) \cdots \mu_{X_0}(x_0) d\nu_{X_0}(x_0)$$

Identification: IP-weighting

We have that:

$$\mathbb{E}_P[Y^{A_0=a_0^*, A_1=a_1^*, \ldots, A_K=a_K^*}]$$

$$= \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} \sum_{y=0,1} y \mu_Y(y \mid \bar{x}_K, \bar{a}_K^*)$$

$$\mu_{X_K}(x_K \mid \bar{x}_{K-1}, \bar{a}_{K-1}^*) d\nu_{X_K}(x_K) \cdots \mu_{X_0}(x_0) d\nu_{X_0}(x_0)$$

$$= \int_{\mathbb{R}^d} \sum_{a_0=0,1} \cdots \int_{\mathbb{R}^d} \sum_{a_K=0,1} \sum_{y=0,1} \frac{\prod_{k=0}^K 1\{a_k = a_k^*\}}{\prod_{k=0}^K \pi_{A_k}(a_k^* \mid \bar{x}_k, \bar{a}_{k-1})\}} y \mu_Y(y \mid \bar{x}_K, \bar{a}_K)$$

$$\pi_{A_K}(a_K \mid \bar{x}_K, \bar{a}_K) \mu_{X_K}(x_K \mid \bar{x}_{K-1}, \bar{a}_{K-1}) d\nu_{X_K}(x_K)$$

$$\cdots \pi_0(a_0 \mid x_0) \mu_{X_0}(x_0) d\nu_{X_0}(x_0)$$

Identification: IP-weighting

We have that:

$$\mathbb{E}_P[Y^{A_0=a_0^*,A_1=a_1^*,\dots,A_K=a_K^*}]$$

$$= \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} \sum_{y=0,1} y\mu_Y(y \mid \bar{x}_K, \bar{a}_K^*)$$

$$\mu_{X_K}(x_K \mid \bar{x}_{K-1}, \bar{a}_{K-1}^*) d\nu_{X_K}(x_K) \cdots \mu_{X_0}(x_0) d\nu_{X_0}(x_0)$$

$$= \int_{\mathbb{R}^d} \sum_{a_0=0,1} \cdots \int_{\mathbb{R}^d} \sum_{a_K=0,1} \sum_{y=0,1} \frac{\prod_{k=0}^K 1\{a_k = a_k^*\}}{\prod_{k=0}^K \pi_{A_k}(a_k^* \mid \bar{x}_k, \bar{a}_{k-1})\}} y\mu_Y(y \mid \bar{x}_K, \bar{a}_K)$$

$$\pi_{A_K}(a_K \mid \bar{x}_K, \bar{a}_K)\mu_{X_K}(x_K \mid \bar{x}_{K-1}, \bar{a}_{K-1}) d\nu_{X_K}(x_K)$$

$$\cdots \pi_0(a_0 \mid x_0)\mu_{X_0}(x_0) d\nu_{X_0}(x_0)$$

$$= \mathbb{E}_P\left[\frac{\prod_{k=0}^K 1\{A_k = a_k^*\}}{\prod_{k=0}^K \pi_{A_k}(a_k^* \mid \bar{X}_k, \bar{A}_{k-1})} Y\right].$$

Identification: IP-weighting

We have that:

$$
\begin{aligned}
&\mathbb{E}_P[Y^{A_0=a_0^*,A_1=a_1^*,\ldots,A_K=a_K^*}] \\
&= \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} \sum_{y=0,1} y \mu_Y(y \mid \bar{x}_K, \bar{a}_K^*) \\
&\qquad\qquad \mu_{X_K}(x_K \mid \bar{x}_{K-1}, \bar{a}_{K-1}^*) d\nu_{X_K}(x_K) \cdots \mu_{X_0}(x_0) d\nu_{X_0}(x_0) \\
&= \int_{\mathbb{R}^d} \sum_{a_0=0,1} \cdots \int_{\mathbb{R}^d} \sum_{a_K=0,1} \sum_{y=0,1} \frac{\prod_{k=0}^K 1\{a_k = a_k^*\}}{\prod_{k=0}^K \pi_{A_k}(a_k^* \mid \bar{x}_k, \bar{a}_{k-1}))\}} y \mu_Y(y \mid \bar{x}_K, \bar{a}_K) \\
&\qquad\qquad \pi_{A_K}(a_K \mid \bar{x}_K, \bar{a}_K) \mu_{X_K}(x_K \mid \bar{x}_{K-1}, \bar{a}_{K-1}) d\nu_{X_K}(x_K) \\
&\qquad\qquad\qquad \cdots \pi_0(a_0 \mid x_0) \mu_{X_0}(x_0) d\nu_{X_0}(x_0) \\
&= \mathbb{E}_P\left[\frac{\prod_{k=0}^K 1\{A_k = a_k^*\}}{\prod_{k=0}^K \pi_{A_k}(a_k^* \mid \bar{X}_k, \bar{A}_{k-1})} Y \right].
\end{aligned}
$$

Identification: g-formula & iterated expectations

The g-formula:

$$
\begin{aligned}
&\mathbb{E}_P\big[Y^{A_0=a_0^*,A_1=a_1^*,\ldots,A_K=a_K^*}\big] \\
&\quad = \int_{\mathbb{R}^d}\cdots\int_{\mathbb{R}^d}\sum_{y=0,1} y\mu_Y(y \mid \bar{x}_K, \bar{a}_K^*) \\
&\qquad\qquad \mu_{X_K}(x_K \mid \bar{x}_{K-1}, \bar{a}_{K-1}^*)d\nu_{X_K}(x_K)\cdots\mu_{X_0}(x_0)d\nu_{X_0}(x_0)
\end{aligned}
$$

can also be written as a sequence of iterated conditional expectations.

Identification: g-formula & iterated expectations

$$
\begin{aligned}
&\mathbb{E}_P\big[Y^{A_0=a_0^*, A_1=a_1^*, \ldots, A_K=a_K^*}\big] \\
&\quad = \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} \sum_{y=0,1} y \mu_Y(y \mid \bar{x}_K, \bar{a}_K^*) \\
&\qquad\qquad \mu_{X_K}(x_K \mid \bar{x}_{K-1}, \bar{a}_{K-1}^*) d\nu_{X_K}(x_K) \cdots \mu_{X_0}(x_0) d\nu_{X_0}(x_0)
\end{aligned}
$$

Identification: g-formula & iterated expectations

$$
\begin{aligned}
&\mathbb{E}_P\left[Y^{A_0=a_0^*, A_1=a_1^*, \dots, A_K=a_K^*}\right] \\
&= \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} \sum_{y=0,1} y \mu_Y(y \mid \bar{x}_K, \bar{a}_K^*) \\
&\qquad \mu_{X_K}(x_K \mid \bar{x}_{K-1}, \bar{a}_{K-1}^*) d\nu_{X_K}(x_K) \cdots \mu_{X_0}(x_0) d\nu_{X_0}(x_0) \\
&= \int_{\mathbb{R}^d} \sum_{a_0=0,1} \cdots \int_{\mathbb{R}^d} \sum_{a_K=0,1} \mathbb{E}[Y \mid \bar{X}_K = \bar{x}_K, \bar{A}_K = \bar{a}_K] \\
&\qquad 1\{a_k = a_k^*\} \mu_{X_K}(x_K \mid \bar{x}_{K-1}, \bar{a}_{K-1}) d\nu_{X_K}(x_K) \\
&\qquad \cdots 1\{a_0 = a_0^*\} \mu_{X_0}(x_0) d\nu_{X_0}(x_0)
\end{aligned}
$$

Identification: g-formula & iterated expectations

$$
\begin{aligned}
&\mathbb{E}_P\left[Y^{A_0=a_0^*,A_1=a_1^*,\dots,A_K=a_K^*}\right] \\
&\quad= \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} \sum_{y=0,1} y\mu_Y(y \mid \bar{x}_K, \bar{a}_K^*) \\
&\qquad\qquad \mu_{X_K}(x_K \mid \bar{x}_{K-1}, \bar{a}_{K-1}^*) d\nu_{X_K}(x_K) \cdots \mu_{X_0}(x_0) d\nu_{X_0}(x_0) \\
&\quad= \int_{\mathbb{R}^d} \sum_{a_0=0,1} \cdots \int_{\mathbb{R}^d} \sum_{a_K=0,1} \mathbb{E}[Y \mid \bar{X}_K = \bar{x}_K, \bar{A}_K = \bar{a}_K] \\
&\qquad\qquad 1\{a_k = a_k^*\} \mu_{X_K}(x_K \mid \bar{x}_{K-1}, \bar{a}_{K-1}) d\nu_{X_K}(x_K) \\
&\qquad\qquad \cdots 1\{a_0 = a_0^*\} \mu_{X_0}(x_0) d\nu_{X_0}(x_0)
\end{aligned}
$$

Identification: g-formula & iterated expectations

$$
\begin{aligned}
&\mathbb{E}_P\big[Y^{A_0=a_0^*,A_1=a_1^*,\dots,A_K=a_K^*}\big] \\
&\quad = \int_{\mathbb{R}^d}\cdots\int_{\mathbb{R}^d}\sum_{y=0,1} y\mu_Y(y \mid \bar{x}_K,\bar{a}_K^*) \\
&\qquad\qquad \mu_{X_K}(x_K \mid \bar{x}_{K-1},\bar{a}_{K-1}^*)d\nu_{X_K}(x_K)\cdots\mu_{X_0}(x_0)d\nu_{X_0}(x_0) \\
&\quad = \int_{\mathbb{R}^d}\sum_{a_0=0,1}\cdots\int_{\mathbb{R}^d}\sum_{a_K=0,1}\mathbb{E}\big[Y \mid \bar{X}_K=\bar{x}_K,\bar{A}_K=\bar{a}_K\big] \\
&\qquad\qquad 1\{a_k=a_k^*\}\mu_{X_K}(x_K \mid \bar{x}_{K-1},\bar{a}_{K-1})d\nu_{X_K}(x_K) \\
&\qquad\qquad \cdots 1\{a_0=a_0^*\}\mu_{X_0}(x_0)d\nu_{X_0}(x_0)
\end{aligned}
$$

Define:

$$\bar{Q}_{K+1}(\bar{x}_K,\bar{a}_K) = \mathbb{E}_P\big[Y \mid \bar{X}_K=\bar{x}_K,\bar{A}_K=\bar{a}_K\big]$$

Identification: g-formula & iterated expectations

$$
\begin{aligned}
&\mathbb{E}_P\big[Y^{A_0=a_0^*, A_1=a_1^*, \dots, A_K=a_K^*}\big] \\
&\quad = \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} \sum_{y=0,1} y \mu_Y(y \mid \bar{x}_K, \bar{a}_K^*) \\
&\qquad\qquad \mu_{X_K}(x_K \mid \bar{x}_{K-1}, \bar{a}_{K-1}^*) d\nu_{X_K}(x_K) \cdots \mu_{X_0}(x_0) d\nu_{X_0}(x_0) \\
&\quad = \int_{\mathbb{R}^d} \sum_{a_0=0,1} \cdots \int_{\mathbb{R}^d} \sum_{a_K=0,1} \bar{Q}_{K+1}(\bar{x}_K, a_K^*, \bar{a}_{K-1}) \\
&\qquad\qquad 1\{a_k = a_k^*\} \mu_{X_K}(x_K \mid \bar{x}_{K-1}, \bar{a}_{K-1}) d\nu_{X_K}(x_K) \\
&\qquad\qquad \cdots 1\{a_0 = a_0^*\} \mu_{X_0}(x_0) d\nu_{X_0}(x_0)
\end{aligned}
$$

Define:

$$\bar{Q}_{K+1}(\bar{x}_K, \bar{a}_K) = \mathbb{E}_P\big[Y \mid \bar{X}_K = \bar{x}_K, \bar{A}_K = \bar{a}_K\big]$$

Identification: g-formula & iterated expectations

$$
\begin{aligned}
&\mathbb{E}_P\big[Y^{A_0=a_0^*,A_1=a_1^*,\dots,A_K=a_K^*}\big] \\
&\quad = \int_{\mathbb{R}^d}\cdots\int_{\mathbb{R}^d}\sum_{y=0,1} y\mu_Y(y \mid \bar{x}_K, \bar{a}_K^*) \\
&\qquad\qquad \mu_{X_K}(x_K \mid \bar{x}_{K-1}, \bar{a}_{K-1}^*)d\nu_{X_K}(x_K)\cdots\mu_{X_0}(x_0)d\nu_{X_0}(x_0) \\
&\quad = \int_{\mathbb{R}^d}\sum_{a_0=0,1}\cdots\int_{\mathbb{R}^d}\bar{Q}_{K+1}(\bar{x}_K, a_K^*, \bar{a}_{K-1}) \\
&\qquad\qquad \mu_{X_K}(x_K \mid \bar{x}_{K-1}, \bar{a}_{K-1})d\nu_{X_K}(x_K) \\
&\qquad\qquad \cdots 1\{a_0 = a_0^*\}\mu_{X_0}(x_0)d\nu_{X_0}(x_0)
\end{aligned}
$$

Define:

$$\bar{Q}_{K+1}(\bar{x}_K, \bar{a}_K) = \mathbb{E}_P\big[Y \mid \bar{X}_K = \bar{x}_K, \bar{A}_K = \bar{a}_K\big]$$

Identification: g-formula & iterated expectations

$$
\begin{aligned}
&\mathbb{E}_P\left[Y^{A_0=a_0^*,A_1=a_1^*,\dots,A_K=a_K^*}\right] \\
&\quad = \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} \sum_{y=0,1} y \mu_Y(y \mid \bar{x}_K, \bar{a}_K^*) \\
&\qquad\qquad \mu_{X_K}(x_K \mid \bar{x}_{K-1}, \bar{a}_{K-1}^*) d\nu_{X_K}(x_K) \cdots \mu_{X_0}(x_0) d\nu_{X_0}(x_0) \\
&\quad = \int_{\mathbb{R}^d} \sum_{a_0=0,1} \cdots \int_{\mathbb{R}^d} \bar{Q}_{K+1}(\bar{x}_K, a_K^*, \bar{a}_{K-1}) \\
&\qquad\qquad \mu_{X_K}(x_K \mid \bar{x}_{K-1}, \bar{a}_{K-1}) d\nu_{X_K}(x_K) \\
&\qquad\qquad \cdots 1\{a_0 = a_0^*\} \mu_{X_0}(x_0) d\nu_{X_0}(x_0)
\end{aligned}
$$

Define:

$$
\bar{Q}_{K+1}(\bar{x}_K, \bar{a}_K) = \mathbb{E}_P\left[Y \mid \bar{X}_K = \bar{x}_K, \bar{A}_K = \bar{a}_K\right]
$$

$$
\bar{Q}_K(\bar{x}_{K-1}, \bar{a}_{K-1}) = \mathbb{E}_P\left[Q_{K+1}(\bar{x}_K, a_K^*, \bar{a}_{K-1}) \mid \bar{X}_{K-1} = \bar{x}_{K-1}, \bar{A}_{K-1} = \bar{a}_{K-1}\right]
$$

Identification: g-formula & iterated expectations

$$
\begin{aligned}
&\mathbb{E}_P[Y^{A_0=a_0^*, A_1=a_1^*, \dots, A_K=a_K^*}] \\
&= \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} \sum_{y=0,1} y \mu_Y(y \mid \bar{x}_K, \bar{a}_K^*) \\
&\qquad \mu_{X_K}(x_K \mid \bar{x}_{K-1}, \bar{a}_{K-1}^*) d\nu_{X_K}(x_K) \cdots \mu_{X_0}(x_0) d\nu_{X_0}(x_0) \\
&= \int_{\mathbb{R}^d} \sum_{a_0=0,1} \cdots \int_{\mathbb{R}^d} \bar{Q}_{K+1}(\bar{x}_K, a_K^*, \bar{a}_{K-1}) \\
&\qquad \mu_{X_K}(x_K \mid \bar{x}_{K-1}, \bar{a}_{K-1}) d\nu_{X_K}(x_K) \\
&\qquad \cdots 1\{a_0 = a_0^*\} \mu_{X_0}(x_0) d\nu_{X_0}(x_0) \\
&= \int_{\mathbb{R}^d} \sum_{a_0=0,1} \cdots \int_{\mathbb{R}^d} \sum_{a_{K-1}=0,1} \bar{Q}_K(\bar{x}_{K-1}, \bar{a}_{K-1}) \\
&\qquad 1\{a_{K-1} = a_{K-1}^*\} \mu_{X_{K-1}}(x_{K-1} \mid \bar{x}_{K-2}, \bar{a}_{K-2}) d\nu_{X_{K-1}}(x_{K-1}) \\
&\qquad \cdots 1\{a_0 = a_0^*\} \mu_{X_0}(x_0) d\nu_{X_0}(x_0)
\end{aligned}
$$

Define:

$$\bar{Q}_{K+1}(\bar{x}_K, \bar{a}_K) = \mathbb{E}_P[Y \mid \bar{X}_K = \bar{x}_K, \bar{A}_K = \bar{a}_K]$$

$$\bar{Q}_K(\bar{x}_{K-1}, \bar{a}_{K-1}) = \mathbb{E}_P[Q_{K+1}(\bar{x}_K, a_K^*, \bar{a}_{K-1}) \mid \bar{X}_{K-1} = \bar{x}_{K-1}, \bar{A}_{K-1} = \bar{a}_{K-1}]$$

Identification: g-formula & iterated expectations

$$\mathbb{E}_P\left[Y^{A_0=a_0^*,A_1=a_1^*,\dots,A_K=a_K^*}\right]$$

$$= \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} \sum_{y=0,1} y\mu_Y(y \mid \bar{x}_K, \bar{a}_K^*)$$
$$\mu_{X_K}(x_K \mid \bar{x}_{K-1}, \bar{a}_{K-1}^*) d\nu_{X_K}(x_K) \cdots \mu_{X_0}(x_0) d\nu_{X_0}(x_0)$$

$$= \int_{\mathbb{R}^d} \sum_{a_0=0,1} \cdots \int_{\mathbb{R}^d} \bar{Q}_{K+1}(\bar{x}_K, a_K^*, \bar{a}_{K-1})$$
$$\mu_{X_K}(x_K \mid \bar{x}_{K-1}, \bar{a}_{K-1}) d\nu_{X_K}(x_K)$$
$$\cdots 1\{a_0 = a_0^*\}\mu_{X_0}(x_0) d\nu_{X_0}(x_0)$$

$$= \int_{\mathbb{R}^d} \sum_{a_0=0,1} \cdots \int_{\mathbb{R}^d} \bar{Q}_{K}(\bar{x}_{K-1}, a_{K-1}^*, \bar{a}_{K-2})$$
$$\mu_{X_{K-1}}(x_{K-1} \mid \bar{x}_{K-2}, \bar{a}_{K-2}) d\nu_{X_{K-1}}(x_{K-1})$$
$$\cdots 1\{a_0 = a_0^*\}\mu_{X_0}(x_0) d\nu_{X_0}(x_0)$$

Define:

$$\bar{Q}_{K+1}(\bar{x}_K, \bar{a}_K) = \mathbb{E}_P\left[Y \mid \bar{X}_K = \bar{x}_K, \bar{A}_K = \bar{a}_K\right]$$
$$\bar{Q}_{K}(\bar{x}_{K-1}, \bar{a}_{K-1}) = \mathbb{E}_P\left[Q_{K+1}(\bar{x}_K, a_K^*, \bar{a}_{K-1}) \mid \bar{X}_{K-1} = \bar{x}_{K-1}, \bar{A}_{K-1} = \bar{a}_{K-1}\right]$$

Identification: g-formula & iterated expectations

$$
\begin{aligned}
&\mathbb{E}_P\big[Y^{A_0=a_0^*, A_1=a_1^*, \dots, A_K=a_K^*}\big] \\
&= \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} \sum_{y=0,1} y \mu_Y(y \mid \bar{x}_K, \bar{a}_K^*) \\
&\qquad \mu_{X_K}(x_K \mid \bar{x}_{K-1}, \bar{a}_{K-1}^*) d\nu_{X_K}(x_K) \cdots \mu_{X_0}(x_0) d\nu_{X_0}(x_0) \\
&= \int_{\mathbb{R}^d} \sum_{a_0=0,1} \cdots \int_{\mathbb{R}^d} \bar{Q}_{K+1}(\bar{x}_K, a_K^*, \bar{a}_{K-1}) \\
&\qquad \mu_{X_K}(x_K \mid \bar{x}_{K-1}, \bar{a}_{K-1}) d\nu_{X_K}(x_K) \\
&\qquad \cdots 1\{a_0 = a_0^*\} \mu_{X_0}(x_0) d\nu_{X_0}(x_0) \\
&= \int_{\mathbb{R}^d} \sum_{a_0=0,1} \cdots \int_{\mathbb{R}^d} \bar{Q}_K(\bar{x}_{K-1}, a_{K-1}^*, \bar{a}_{K-2}) \\
&\qquad \mu_{X_{K-1}}(x_{K-1} \mid \bar{x}_{K-2}, \bar{a}_{K-2}) d\nu_{X_{K-1}}(x_{K-1}) \\
&\qquad \cdots 1\{a_0 = a_0^*\} \mu_{X_0}(x_0) d\nu_{X_0}(x_0)
\end{aligned}
$$

Define:

$$
\bar{Q}_{K+1}(\bar{x}_K, \bar{a}_K) = \mathbb{E}_P\big[Y \mid \bar{X}_K = \bar{x}_K, \bar{A}_K = \bar{a}_K\big]
$$

$$
\bar{Q}_K(\bar{x}_{K-1}, \bar{a}_{K-1}) = \mathbb{E}_P\big[Q_{K+1}(\bar{x}_K, a_K^*, \bar{a}_{K-1}) \mid \bar{X}_{K-1} = \bar{x}_{K-1}, \bar{A}_{K-1} = \bar{a}_{K-1}\big]
$$

Identification: g-formula & iterated expectations

Full steps that represent $\mathbb{E}_P[Y^{A_0=a_0^*,A_1=a_1^*,\dots,A_K=a_K^*}]$ as a sequence of iterated conditional expectations:

$$\bar{Q}_{K+1}(\bar{x}_K, \bar{a}_K) = \mathbb{E}_P[Y \mid \bar{X}_K = \bar{x}_K, \bar{A}_K = \bar{a}_K]$$

$$\bar{Q}_K(\bar{x}_{K-1}, \bar{a}_{K-1}) = \mathbb{E}_P[Q_{K+1}(\bar{X}_K, a_K^*, \bar{A}_{K-1}) \mid \bar{X}_{K-1} = \bar{x}_{K-1}, \bar{A}_{K-1} = \bar{a}_{K-1}]$$

$$\vdots$$

$$\bar{Q}_k(\bar{x}_{k-1}, \bar{a}_{k-1}) = \mathbb{E}_P[Q_{k+1}(\bar{X}_k, a_k^*, \bar{A}_{k-1}) \mid \bar{X}_{k-1} = \bar{x}_{k-1}, \bar{A}_{k-1} = \bar{a}_{k-1}]$$

$$\vdots$$

$$\bar{Q}_2(\bar{x}_1, \bar{a}_1) = \mathbb{E}_P[Q_3(\bar{X}_2, a_2^*, \bar{A}_1) \mid \bar{X}_1 = \bar{x}_1, \bar{A}_1 = \bar{a}_1]$$

$$\bar{Q}_1(x_0, a_0) = \mathbb{E}_P[Q_2(\bar{X}_1, a_1^*, A_0) \mid X_0 = x_0, A_0 = a_0]$$

$$\mathbb{E}_P[Y^{A_0=a_0^*,A_1=a_1^*,\dots,A_K=a_K^*}] = \mathbb{E}_P[\bar{Q}_1(x_0, a_0^*)].$$

Identification (summary)

1. IP-weighting:

$$\mathbb{E}_P\big[Y^{A_0=a_0^*,A_1=a_1^*,\dots,A_K=a_K^*}\big] = \mathbb{E}_P\left[\frac{\prod_{k=0}^K 1\{A_k = a_k^*\}}{\prod_{k=0}^K \pi_{A_k}(a_k^* \mid \bar{X}_k, \bar{A}_{k-1})} Y\right]$$

2. Sequence of iterated conditional expectations:

$$\bar{Q}_{K+1}(\bar{x}_K, \bar{a}_K) = \mathbb{E}_P\big[Y \mid \bar{X}_K = \bar{x}_K, \bar{A}_K = \bar{a}_K\big]$$

and iteratively for $k = K, K-1, \dots, 1$,

$$\bar{Q}_k(\bar{x}_{k-1}, \bar{a}_{k-1}) = \mathbb{E}_P\big[\bar{Q}_{k+1}(\bar{X}_k, a_k^*, \bar{A}_{k-1}) \mid \bar{X}_{k-1} = \bar{x}_{k-1}, \bar{A}_{k-1} = \bar{a}_{k-1}\big]$$

so that

$$\mathbb{E}_P\big[Y^{A_0=a_0^*,A_1=a_1^*,\dots,A_K=a_K^*}\big] = \mathbb{E}_P\big[\bar{Q}_1(X_0, a_0^*)\big].$$

Practical 1: Kreif et al. (2017) as an example

In this practical we consider the study by Kreif et al. as an example:

- Data structure, static and dynamic intervention, time-dependent treatment interventions.
- IP-weighting, g-formula, TMLE.

Questions for the paper that you should go over can be found in: **day3_practical1.pdf**.