Causal Inference

Chapter 5. Studies with treatment at multiple time points:
the case of sequentially ignorable assignment.

1 Introduction: what type of studies we consider.

(1) Treatments with multiple time points, where those treatments assignment is ignorable conditionally on the observed history.

If we can justify the above assumption, this is a possible template for

- experiments;
- observational studies, e.g., with
  - ⋆ patients visiting doctors at different times.
  - ⋆ workers exposed to hazards at the workplace (related to health worker survivor effect.

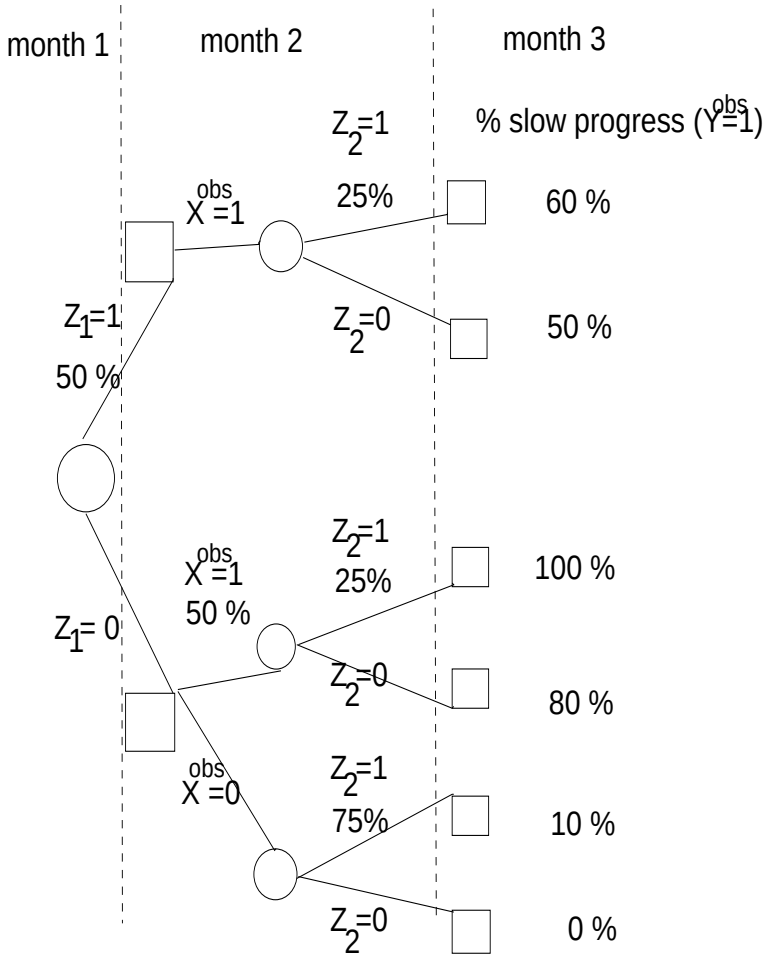
(2) Example and notation.

month	action	potential outcome	observed value
1	give treatment z_1 (1=high)		$Z_{i,1}$
2	(i) measure cancer progression		X_i^{obs}
	(ii) give treatment z_2 (1=high)		$Z_{i,2}$
3	measure cancer progression	$Y_i(z_1, z_2)$	Y_i^{obs}

Context (also see figure 1).

- Patients with cancer.
- Visiting doctor at two time points t_1,t_2.
- $z_t(t = 1, 2)$: possible treatment at time t; $Z_{i,t}$: the observed treatment at time t.
- X_i^{obs}: observed cancer progression at time 2.
- $Y_i(z_1, z_2)$: potential outcome at time 3.
- $Y_i^{obs} = Y_i(Z_{i,1}, Z_{i,2})$: observed outcomes.

Figure 1: Example with treatment at two time points and sequentially ignorable assignment



(3) A goal.

Compare cancer progression Y (1 = slow) between:

- taking always high dose, $z_1 = z_2 = 1$, that is, pr$(Y(1,1) = 1)$, and
- taking always low dose, $z_1 = z_2 = 0$, that is, pr$(Y(0,0) = 1)$.

Note 1. We only control (z_1, z_2), that is why the potential outcomes $Y_i()$ are only a function of z_1, z_2 and not also of X.

Note 2. That does not mean we are only interested in marginal effects.

(4) Assumption of sequential ignorability (Robins, 1986).

Sequential ignorability : treatment at time t is randomized with probabilities depending on the observed past, *including covariates, outcomes*.

In terms of the Rubin-Causal-Model, at any time t :

$$\{Y_i(z_1, z_2), \text{all } z_1, z_2\} \coprod Z_{i,t} \mid \text{Obs-Info}(t)$$

Here, Obs-Info(1)=nothing, and Obs-Info(2) $= (Z_{i,1}, X_i^{obs})$.

2 Problems with standard methods.

Recall goal: to estimate pr$(Y(0,0) = 1)$, pr$(Y(1,1) = 1)$.

Approach 1. "Do not condition on progression X^{obs} because it is an outcome":

$$\underset{60\%}{\text{pr}(Y^{obs} = 1 | Z_1 = 1, Z_2 = 1)} - \underset{60\%}{\text{pr}(Y^{obs} = 1 | Z_1 = 0, Z_2 = 0)} = 0\%$$

Approach 2. "Condition on intermediate progression X^{obs} because it was used in deciding treatment Z_2":

$$\underset{60\%}{\text{pr}(Y^{obs} = 1 | Z_1 = 1, Z_2 = 1, X^{obs} = 1)} - \underset{80\%}{\text{pr}(Y^{obs} = 1 | Z_1 = 0, Z_2 = 0, X^{obs} = 1)} = -20\%$$

Note. Adjusting and not adjusting for the intermediate variable X_i^{obs} in the standard ways is incorrect for the goal (to estimate pr$(Y(0,0) = 1)$, pr$(Y(1,1) = 1)$).

3 Identifiability of causal effects under the assumption of sequential ignorability.

For sequential ignorability: g-computation, (Robins, 1986)
(induction on result of Rosenbaum and Rubin (1983)):

$$\{Y_i(z_1, z_2), \text{all } z_1, z_2\} \coprod Z_{i,t} \mid \text{Observed info until time } t$$

By assumption of sequential ignorability, we have:

$$\begin{aligned}
\text{pr}(Y(0,0)=1) &= \text{pr}(Y(0,0)=1|Z_1=0) \quad \text{(ignorability at t=1)} \\
\text{(Law of total probability)} \quad &= \sum_{X^{obs}=0,1} \text{pr}(Y(0,0)=1|Z_1=0,X^{obs})\text{pr}(X^{obs}|Z_1=0) \\
\text{(ignorability at t=2)} \quad &= \sum_{X^{obs}=0,1} \text{pr}(Y(0,0)=1|Z_1=0,X^{obs},Z_2=0)\text{pr}(X^{obs}|Z_1=0) \\
\text{(potnl. outcomes)} \quad &= \sum_{X^{obs}=0,1} \text{pr}(Y^{obs}=1|Z_1=0,X^{obs},Z_2=0)\text{pr}(X^{obs}|Z_1=0) \\
&= 0\%(50\%) + 80\%(50\%) = 40\%
\end{aligned}$$

$\text{pr}(Y(1,1)=1) = 60\%$ so, causal effect $\text{pr}(Y(1,1)=1) - \text{pr}(Y(0,0)=1) = \quad +20\%$

We can also estimate $pr(Y_i(0,1)=1)$ and $pr(Y_i(1,0))=1$ using similar arguments as above (omitted here). And we have $\text{pr}(Y(0,1)=1) = 55\%$, and $\text{pr}(Y(1,0)=1) = 50\%$.

4 Modelling the outcome partly, and modelling the propensity score of assignment at different times, $\text{pr}(Z_{i,1})$, and $\text{pr}(Z_{i,2} \mid X_i^{obs}, Z_{i,1})$.

We can model

(i) $\text{pr}(Y_i(z_1,z_2)=y \mid \alpha_{(y)}) = h(y,z_1,z_2,\alpha_{(y)})$;

(ii) $\text{pr}(Z_{i,1}=1) = \pi_1$;

(iii) $\text{pr}(Z_{i,2}=1 \mid X_i^{obs}=x, Z_{i,1}=z_1, \alpha_{(z)}) = \pi_2(x,z_1,\alpha_{(z)})$,

where define $\alpha^0_{(y)}, \alpha^0_{(z)}$ to be the true value.

Idea. Generalize Horwitz-Thompson estimator to multi-stage.

Define the score to be $S(y,z_1,z_2,\alpha) = \frac{\partial}{\partial\alpha}\log h(y,z_1,z_2,\alpha_{(y)})$.

Then we consider expression

$$\sum_i S(Y_i^{obs}, Z_{i,1}, Z_{i,2}, \alpha_{(y)})W_i = 0, \qquad (ex.1)$$

where $W_i^{-1} = \pi_1^{Z_{i,1}}(1-\pi_1)^{1-Z_{i,1}}(\pi_2(X^{obs},Z_{i,1},\alpha_{(z)}))^{Z_{i,2}}(1-\pi_2(X^{obs},Z_{i,1},\alpha_{(z)}))^{1-Z_{i,2}}$.

Note 1. Solving (ex.1) can give consistent estimates of the model parameters and thus of the causal effect (Robins, Hernan and Brumback, 2000).

Note 2. These estimates, as with Horwitz-Thompson estimator, can have large variance. They can somehow be improved.