STANDARDIZATION AND G-FORMULA

Group meeting:

Causal Inference What If reading

Keling Wang

01 May 2024

STANDARDIZATION

- Basic idea: for a time-fixed setting, given Y, A, L, and identifiability conditions hold; then the marginal counterfactual outcome

$$E[Y^a] = \sum_l E[Y|A = a, L = l]\Pr[L = l]$$

or

$$\int E[Y|A = a, L = l]f_L(l)\, \mathrm{d}l$$

- **Rationale** and link to Chapter 2 (Conditionally randomized trial setting): The treatment A is randomized with respect to the probability $\Pr[L = l]$ or density $f_L(l)$.

RATIONALE 2

- IPTW and Standardization deal confounders in different ways:
- **IPTW:** create a pseudo-population in which the confounders are balanced. Require estimating treatment probability

$$\Pr[A = a | L = l] \text{ ; or } f_A(a|l)$$

for each treatment level.

- **Standardization:** create a trial in which the treatment is assigned w.r.t. the joint distribution of confounders. Require estimating conditional mean of the outcome

$$E[Y | A = a, L = l]$$

for each individual in the study population.

IDENTIFIABILITY CONDITIONS

- Consistency: $Y = Y^a$ among those whose $A = a$

- (Structural) positivity: $f_A(a|l) > 0$ where $f_L(l) > 0$

- Conditional exchangeability: $Y^a \amalg A | L$

- Note: structural and random violation of positivity for variable space containing continuous random variables.

SIMPLE ESTIMATION

1. In time-fixed settings, only conditional mean part is needed to be estimated. Empirical distribution of L will be used in the second part of the g-formula.

2. Nonparametric estimation: See section 2 example, for single discrete L.

3. Parametric estimation: Require parametrization of the conditional mean part.

ILLUSTRATION EXAMPLE

- 4 steps: Expansion of dataset, outcome modelling, prediction, and standardization by averaging

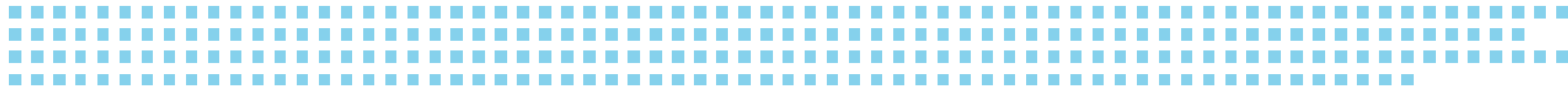- Parametrization choices: be careful about model assumptions

Table 2.2

	L	A	Y
Rheia	0	0	0
Kronos	0	0	1
Demeter	0	0	0
Hades	0	0	0
Hestia	0	1	0
Poseidon	0	1	0
Hera	0	1	0
Zeus	0	1	1
Artemis	1	0	1
Apollo	1	0	1
Leto	1	0	0
Ares	1	1	1
Athena	1	1	1
Hephaestus	1	1	1
Aphrodite	1	1	1
Polyphemus	1	1	1
Persephone	1	1	1
Hermes	1	1	0
Hebe	1	1	0
Dionysus	1	1	0

Second block: All untreated

	L	A	Y
Rheia	0	0	.
Kronos	0	0	.
Demeter	0	0	.
Hades	0	0	.
Hestia	0	0	.
Poseidon	0	0	.
Hera	0	0	.
Zeus	0	0	.
Artemis	1	0	.
Apollo	1	0	.
Leto	1	0	.
Ares	1	0	.
Athena	1	0	.
Hephaestus	1	0	.
Aphrodite	1	0	.
Polyphemus	1	0	.
Persephone	1	0	.
Hermes	1	0	.
Hebe	1	0	.
Dionysus	1	0	.

Third block: All treated

|	L	A	Y
Rheia	0	1	.
Kronos	0	1	.
Demeter	0	1	.
Hades	0	1	.
Hestia	0	1	.
Poseidon	0	1	.
Hera	0	1	.
Zeus	0	1	.
Artemis	1	1	.
Apollo	1	1	.
Leto	1	1	.
Ares	1	1	.
Athena	1	1	.
Hephaestus	1	1	.
Aphrodite	1	1	.
Polyphemus	1	1	.
Persephone	1	1	.
Hermes	1	1	.
Hebe	1	1	.
Dionysus	1	1	.

NOTE

- Difference between standardization and outcome regression:

    Standardization: marginal mean

    Outcome regression: conditional mean (conditioning on L).

SOME OTHER TOPICS
About g-formula

- 1. Big g-formula

- 2. front door formula

- 3. g-formula in time-varying settings

- 4. Estimation choices
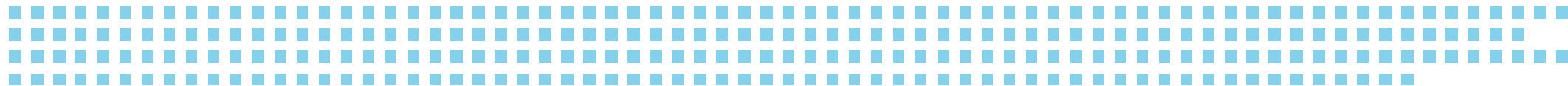
QUESTIONS?

And suggestions?

REFERENCES