Estimating Treatment Effects

Daniel Tompsett

Great Ormond Street Institute of Child Health, University College London

d.tompsett@ucl.ac.uk

Work funded by MRC Methodology Grant: MR/R025215/1

One-day Course on Target Trial Emulation

28th November 2022

G-methods

The three g-methods are broadly speaking

1. Inverse Probability of Treatment weighting (IPTW) of Marginal Structural Models (MSMs).
2. G-computation, a parametric form of the g-formula methods.
3. G-estimation of Marginal Structural Mean Models (MSMMs).

In the following slides we outline the necessary theory and assumptions for g-methods.

Inverse Propensity of Treatment Weighting

- Inverse Probability of Treatment Weighting (IPTW) is the most popular (and arguably the simplest) of the g-methods.
- It seeks to control for confounders by creating a pseudo-population in which the arrows between the exposure and confounders are removed.
- In practice this amounts to weighting the outcome model by the inverse of the probability of being treated, conditional on all measured confounders.
- IPTW is typically used to fit a causal analysis outcome model known as a Marginal Structural Model (MSM).

IPTW: MSM for Scenario A

- A Marginal Structural Model (MSM) models the counterfactual outcome Y^a as a simple function of a.
- For a time-fixed exposure (Scenario A), the model takes the form

$$E[Y^a] = \beta_0 + \beta_1 a$$

- Parameterised in this way, we have that

$$E[Y^1] - E[Y^0] = \beta_1$$

- Thus β_1 automatically gives us the average treatment effect.
- Note: If Y is Binary, the MSM can be modeled as a risk ratio instead.

IPTW: ITT, Scenario A

- Let $A_1 = A$ and $L_1 = L$. IPTW weights for baseline assignment are calculated for each individual j as

$$W_j = \frac{1}{P(A_j|L_j)}$$

- It is not uncommon for the W_j to exhibit extreme values. Hence it is common to use stabilised weights instead

$$W_j = \frac{P(A_j)}{P(A_j|L_j)}$$

- The MSM can then be fitted via a simple regression of Y on A, weighted by W_j, to get the ITT.

IPTW: ITT, Scenario A

- To calculate $P(A_j|L_j)$, we fit a regression model for the probability of being treated, given the confounders.

$$E(P(A = 1|L)) = \delta_0 + \delta_1 L$$

- Predicted values from these model for each individual j are then calculated as $ps_j = p_{A=1|L_j}$
- We then generate the probabilities of the treatment actual received, p_j. These are:
  - If an individual was treated ($A = 1$), $p_j = ps_j$
  - If an individual was not treated ($A = 0$), $p_j = 1 - ps_j$
- Weights can then be calculated from the inverse of these probabilities, and the MSM fitted.

IPTW: Scenario B

- With a time varying treatment with T time periods, a possible specification of an MSM is

$$E[Y^{\bar{a}}] = \beta_0 + \beta_1 \sum_{t=1}^{T} a_t$$

where $\bar{a} = (a_1, \ldots, a_T)$ is a persons history of treatment.

- Alternatively we can parameterise the MSM as.

$$E[Y^{\bar{a}}] = \beta_0 + \sum_{t=1}^{T} \beta_t a_t.$$

- The ATE is then calculated as

$$E[Y^{(1,\ldots,1)}] - E[Y^{(0,\ldots,0)}] = \sum_{t=1}^{T} \beta_t a_t$$

- Thus the sum of the β_t automatically gives us the average

IPTW: PP, Scenario B

- For PP we first define a variable C_t indicating if individuals complied with treatment up to time t ($C_t = 1$).
- We then create inverse censoring V weights based on C, similar to IPW weights.

$$V_j = \frac{I(C_{jT} = 1)}{\prod_{t=1}^{T} P(C_{jt}|\bar{A}_{jt}, \bar{L}_{jt})}$$

- These are estimated using a regression model of C on A_t and any other relevant predictor of compliance.
- The MSM can then be fitted using a regression of Y on A_t, weighted by the **product** of the IPTW and compliance weights.

G-Computation

- For the ITT effect where we only consider baseline assignment, G-computation identifies the potential outcomes (under the usual assumption) as

$$E[Y^a] = \sum_l E[Y|A = a, L = l]P(L = l)$$

- In more complex settings, g-computation is performed by MC simulation by predicting the average value of the outcome under treatment or no treatment interventions by simulations based on a series of parametric models.
- In practice, this can be performed as follows:

G-Comp: ITT, Scenario A

- Let $A_1 = A$ as before. We define a general model for Y given treatment and confounders.
- This is often called the Q-model. This would fit Y against A and L.
- Using the Q-model, take predictions of Y when we replace A with $A = 1$ for all individuals. This obtains predictions of Y^1.

G-Com: ITT, Scenario A and B

- Repeat this replacing $A = 0$ to predict Y^0
- The ATE (in this instance the ITT effect) can then be taken as the average difference in Y^1 and Y^0
- To account for simulation error when estimating Y^1 and Y^0, we can apply the technique for a number of resampled versions of the original dataset.
- This is a type of Monte Carlo Simulation. The ATE can be taken as the average value of the ATE over the simulated data.

G-Comp: Scenario B

- To account for time-varying confounding we must also simulate their values under different interventions.
- We therefore need two models, The Q model and a model for L_2 on A_1 and other relevant variables, *etc.* for later time points.
- Starting with L_2, we predict it using a model for L2 setting $A_1 = 0$ or $A_1 = 1$ for all individuals.
- Call these predicted values $L_2^{pred(1)}$ and $L_2^{pred(0)}$.
- Now predict $Y^{1,1}$ from the fitted Q model, by setting $A_1 = 1$, $A_2 = 1$, and $L_2 = L_2^{pred(1)}$.
- Repeat for $Y^{0,0}$ and estimate the ATE.

Structural Nested Mean Model

- We move on to g-estimation, a more involved causal analysis method. As such we will focus on a single time point.
- G-estimation typically fits a Structural Nested Mean Model (SNMM), which models causal effects as follows.

$$E(Y^1|L) - E(Y^0|L) = f(\beta, a)$$

where $f(\beta, a)$ is some function that defines how a causally affects Y through β.

- $f(\beta, a)$ is sometimes called the "blip" function, its typical form is

$$f(\beta, a) = \beta a$$

Structural Nested Mean Model

There are some key differences from a MSM

1. The causal effect is conditional. It is the effect of treatment (or randomisation to treatment) on outcome, conditional on $L = l$.
2. There is no intercept term in the model. That is we make no parametric assumptions about $E(Y^0|L, A = 0)$.
3. MSMMs can allow the causal effect to depend on L (known as effect modification). For example we can set

$$f(\beta, a) = \beta_1 a + \beta_2 al$$

4. One issue, there is not always a clear means to estimate the ATE. However if Y is continuous and there no effect modification. β can be taken as the ATE.

G-estimation: Scenario A

- G-estimation fits a MSMM taking advantage of the conditional exchangeability assumption

$$Y^a \perp\!\!\!\perp A|L$$

- We posit this, suppose we knew β, and that the SNMM was correct. Then we could estimate the counterfactual outcome as

$$H(a) \equiv Y^0 = Y - f(\beta, a) = Y - a\beta$$

- Then, because of exchangeability, $H(a)$ and A are independent given L, that is in a model

$$E(Y) = \alpha_0 + \alpha_1 H(a) + \alpha_2 L$$

α_1 would be zero. This can be used to search for β.

G-estimation: Scenario A

- In practice, this search can be done by fitting two regression models.
- Firstly we fit a propensity score model for A given its dependencies

$$E(P(A = 1|L)) = \delta_0 + \delta_1 L$$

and take predicted values from this model as the propensity of being treated p_a

- Note that with propensity score, we do NOT need to set $p_a = 1 - p_a$ if $A = 0$.

G-estimation: Scenario A

- Secondly we fit an outcome model for Y, that includes an interaction between p_A and a

$$E(Y|A, L) = \alpha_0 + \alpha_1 L + \beta A + \gamma p_a A$$

- It can be shown that β is the causal effect in the SNMM, the g-estimate.
- This method is the two stage g-estimation process.
- Note we let $f(\beta, a) = \beta a$. For a general $f(\beta, a)$, the parameters β and γ need to have the same structure as $f(\beta, a)$.

Remarks

- Pro: This method is doubly robust, if **either** The propensity, or outcome model are correct, then the result is unbiased.
- Pro: Its very efficient with typically smaller standard errors than other g-methods
- Pro: We have created an R package *gesttools* that will perform two stage g-estimation for a wide variety of situations. Please contact us if you are interested further.
- Con: With multiple treatment periods,two stage g-estimation is challenging theoretically. As such, we don't show the method here.

Conclusion

- Causal analysis and g-methods are a well studied field.
- Here we provided a brief overview of the methodology.
- Extensions to calculation of standard errors via bootstrap, dynamic treatment regimes, and adherence and censoring weights were omitted for simplicity.
- For those interested further we are happy to be contacted, and to provide further resources.
- Thank you!