From Efficient Counterfactual Mean Estimation to Estimating Marginal Structural Models

Jacob M. Chen†, Ilya Shpitser†
jchen459@jhu.edu, ilyas@cs.jhu.edu
†Dept. of Computer Science, Johns Hopkins University

JOHNS HOPKINS UNIVERSITY

Motivation

- Marginal structural models (MSMs) are a class of causal models that assume a parametric model for the counterfactual mean. They are widely used for their interpretability and generalizability to continuous treatments and multiple treatments in longitudinal settings [1].

$$\mathbb{E}[Y(a)] = g(a;\theta)$$

- The goal of the work we present is to estimate θ efficiently.

Example: The Backdoor Model

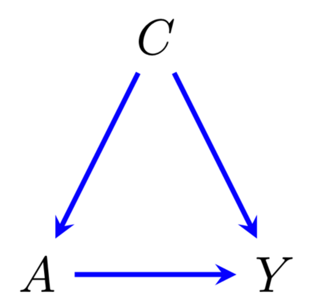

Figure 1: Backdoor graph.

In the backdoor graph, the counterfactual mean is identified as

$$\mathbb{E}[Y(a)] = \sum_C \mathbb{E}[Y \mid a, C]p(C)$$

The backdoor graph can be factorized as

$$p(Y, A, C) = p(Y \mid A, C)\, p(A \mid C)\, p(C).$$

Let

$$q(Y, A, C) = p(Y \mid A, C)\, p^*(A)\, p(C).$$

This is a hypothetical distribution where treatment behaves as if it were randomized.

Suppose we use the squared loss function. Then, we can solve for θ by minimizing

$$\mathbb{E}_q[(Y - g(A;\theta))^2]$$

The subscript q indicates that we take the expectation over the hypothetical randomized distribution. Assuming sufficient regularity to exchange derivative and integral, we take the derivative of above and set it equal to 0 to solve for the minimum:

$$\mathbb{E}_q[h(A)(Y - g(A;\theta))] = 0 \qquad (1)$$

The function h is the derivative of g with respect to θ.
In general, we just need the derivative of the loss function to be (1). For instance, in a logistic model, we can use the cross-entropy loss function [2]. We assume that the derivative of the loss follows (1).

Exercise for audience: The estimating equation (1) can be rewritten as

$$\mathbb{E}\Big[\frac{p^*(A)}{p(A \mid C)}h(A)(Y - g(A;\theta))\Big] = 0$$

This is the canonical inverse-probability weighted estimator for MSMs [1, 2, 3]. A researcher can implement this estimator using standard statistical software with a reweighted regression.

Background: Influence Function Based Estimators

- Influence functions (IFs) measure the "influence" of slightly perturbing the underlying data-generating distribution on an estimand.
- The influence function is defined as the limit, as the amount of perturbation goes to 0, of the difference between the estimand evaluated at the perturbed distribution and the estimand evaluated at the true distribution divided by the amount of perturbation.
- Let ψ be the estimand of interest, F_ϵ be the perturbed distribution, and F be the true distribution.

$$IF = \lim_{\epsilon \to 0} \frac{\psi(F_\epsilon) - \psi(F)}{\epsilon} = \frac{\partial}{\partial \epsilon}\psi(F_\epsilon)\Big|_{\epsilon=0}$$

- Influence function-based estimators avoid **first-order bias**; the estimator is efficient when the **product** (rather than the sum) of the rates of convergence of nuisance functions is efficient. Influence function-based estimators also achieve the nonparametric efficiency bound [4, 5].
- The influence function-based estimator for $\psi = \mathrm{E}[Y(a)]$ in the backdoor case is the canonical doubly robust estimator [4, 5]:

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\mathbb{I}(A_i = a)}{p(A_i \mid C_i)}(Y_i - \mathbb{E}[Y_i \mid a, C_i]) + \mathbb{E}[Y_i \mid a, C_i]$$

References

[1] Hernán, Miguel Ángel, Babette Brumback, and James M. Robins. "Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men." *Epidemiology* 11.5 (2000): 561-570.
[2] Martin, Axel, Michele Santacatterina, and Iván Díaz. "Non-parametric efficient estimation of marginal structural models with multi-valued time-varying treatments." *arXiv preprint arXiv:2409.18782* (2024).
[3] van der Laan, Mark J., and James M. Robins. "Unified approach for causal inference and censored data." *Unified methods for censored longitudinal data and causality*. New York, NY: Springer New York, 2003. 311-370.
[4] Kennedy, Edward H. "Semiparametric doubly robust targeted double machine learning: a review." *Handbook of statistical methods for precision medicine* (2024): 207-236.
[5] Kang, Hyunseung. "Causal Inference: Influence Functions and von Mises Calculus." (2024).
[6] Fulcher, Isabel R., et al. "Robust inference on population indirect causal effects: the generalized front door criterion." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 82.1 (2020): 199-214.
[7] Shpitser, Ilya, and Eric Tchetgen Tchetgen. "Causal inference with a graphical hierarchy of interventions." *Annals of statistics* 44.6 (2016): 2433.

Example: The Backdoor Model (continued)

We can rewrite our estimating equation as:

$$\mathbb{E}_q[h(A)(Y - g(A;\theta))]$$
$$= \sum_A h(A)p^*(A)\mathbb{E}[Y(A)] - \sum_A h(A)p^*(A)g(A;\theta) = 0 \qquad (2)$$

Note that the second term does not depend on the data. Thus, to estimate θ, we only need to estimate the first term.

Proposition 1. Suppose the influence function for E[Y(a)] is known. Then, the influence function for $\sum_A h(A)p^*(A)E[Y(A)]$ is

$$\frac{\partial}{\partial \epsilon}\sum_A p^*(A)h(A)\mathbb{E}[Y(A)]$$
$$= \sum_A p^*(A)h(A)\frac{\partial}{\partial \epsilon}\mathbb{E}[Y(A)]$$

In the case of a continuous treatment, replace the summations with integrals and probabilities with densities, and we still get the correct influence function.

Influence Function of MSM Parameters

To get the influence function for the parameters θ, we use the chain rule. Let $\psi = \sum_A h(A)p^*(A)g(A;\theta)$, where g(A; θ) is the MSM for the counterfactual mean.

$$\frac{\partial \psi}{\partial \epsilon} = IF(\psi)$$
$$\frac{\partial \psi}{\partial \theta}\frac{\partial \theta}{\partial \epsilon} = IF(\psi) \qquad \text{(chain rule)}$$
$$\frac{\partial \theta}{\partial \epsilon} = \Big(\frac{\partial \psi}{\partial \theta}\Big)^{-1} IF(\psi) \qquad (3)$$

$$\frac{\partial \psi}{\partial \theta} = \frac{\partial}{\partial \theta}\sum_A p^*(A)h(A)g(A;\theta)$$
$$= -\mathbb{E}_q\Big[\frac{\partial}{\partial \theta}h(A)(Y - g(A;\theta))\Big]$$

In the backdoor case, this gives us the final desired result

$$\frac{\partial \theta}{\partial \epsilon} = \Big(-\mathbb{E}_q\Big[\frac{\partial}{\partial \theta}h(A)(Y - g(A;\theta))\Big]\Big)^{-1}$$
$$\Big(\frac{p^*(A)}{p(A \mid C)}h(A)(Y - \mathbb{E}[Y \mid A, C]) + \sum_{\tilde{A}} p^*(\tilde{A})h(\tilde{A})\mathbb{E}[Y \mid \tilde{A}, C] - \psi\Big)$$

And the estimating equation

$$\Big[\frac{1}{n}\sum_{i=1}^{n}\frac{p^*(A_i)}{p(A_i \mid C_i)}h(A_i)(Y_i - \mathbb{E}[Y_i \mid A_i, C_i]) +$$
$$\sum_A p^*(A)h(A)\mathbb{E}[Y_i \mid A, C_i]\Big] - \sum_A p^*(A)h(A)g(A;\theta) = 0$$

*This matches the doubly robust MSM estimating equation proposed by Robins and van der Laan [3].

Main Results

Proposition 2. Suppose the true parameters of the MSM (θ^*) are unique. The estimating equation (2) is satisfied at θ^* if and only if the estimator for E[Y(a)] is unbiased.

Note that the MSM estimating equation (2) inherits any "weak double robustness" properties of the counterfactual mean estimator.

Theorem 1. If the influence function of E[Y(a)] is known under discrete treatment, then the influence function for the parameters θ of the MSM is given by (3).

Theorem 2. Suppose we use a root finding algorithm that converges to the true values of the MSM parameters. If the estimator for $\psi = \sum_A h(A)p^*(A)E[Y(A)]$ is asymptotically normal, then (2) is asymptotically normal in estimating the parameters of the MSM.

Note that influence function-based estimators are often provably asymptotically normal under mild conditions, making them good candidates for estimating ψ.

Extension: The Frontdoor MSM

Using the influence function for E[Y(a)] in the frontdoor model given by Fulcher, et al. [6], the estimating equation for the frontdoor MSM parameters is

$$\Big[\frac{1}{n}\sum_{i=1}^{n}\sum_A p^*(A)h(A)\frac{p(M_i \mid A, C_i)}{p(M_i \mid A_i, C_i)}(Y_i - \mathbb{E}[Y_i \mid A_i, M_i, C_i])$$
$$+ \frac{p^*(A_i)}{p(A_i \mid C_i)}h(A_i)\Big(\sum_A \mathbb{E}[Y_i \mid A, M_i, C_i]p(A \mid C_i)$$
$$- \sum_{A,M}\mathbb{E}[Y_i \mid A, M, C_i]p(M \mid A_i, C_i)p(A \mid C_i)\Big)$$
$$+ \sum_{A,M} p^*(A)h(A)\mathbb{E}[Y \mid A_i, M, C_i]p(M \mid A, C_i)\Big]$$
$$- \sum_A p^*(A)h(A)g(A;\theta) = 0$$

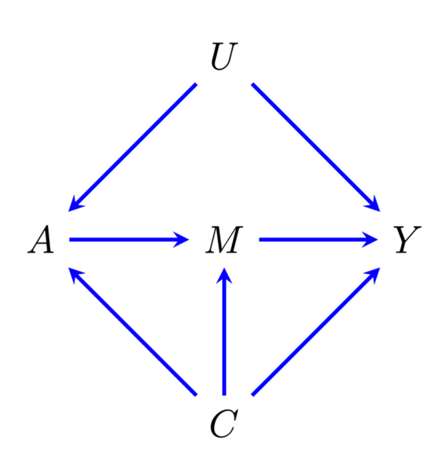

Figure 2: Frontdoor graph.

Future Work

We are investigating influence functions for MSM parameters when the counterfactual mean is identified as a product of conditional probabilities (g-functional), allowing for easier calculation of the influence function for $\psi = \sum_A p^*(A)h(A)E[Y(A)]$ [7].