A Generative Adversarial Framework for Bounding Confounded Causal Effects (Supplementary File)

Yaowei Hu,[1] Yongkai Wu, [2] Lu Zhang, [1] Xintao Wu [1]
[1] University of Arkansas
[2] Clemson University
yaoweihu@uark.edu, yongkaw@clemson.edu, lz006@uark.edu, xintaowu@uark.edu

Pseudocode of ACE Bounding Procedure

The pseudocode of the ACE bounding procedure described in the main paper is given in Algorithm 1.

Algorithm 1: ACE Bounding Procedure

Input : (Confounded) causal graph, variable A with two values a_0, a_1, and variable B.
Output: Lower (or upper) bound of the ACE of A on B.

1. Decompose the causal graph to c-components, and define noise terms $\mathbf{z}$, one for each c-component;
2. For each variable V, define a neural net $v = h_V(\mathsf{pa}_V; \theta_V)$ with parameter θ_V (if V has no parent, the neural net becomes a trivial mapping $v = \theta_V$);
3. For each θ_V, define a generator $\theta_V = G_V(\mathbf{z}_V)$ where $\mathbf{z}_V$ is the noise term of the corresponding c-component;
4. Sample generated distribution $G(\mathbf{z})$, feed together with the observational data into discrimination D to obtain value function $V(G, D)$;
5. Intervene A to a_0 and a_1, sample generated values of B to compute $\mathrm{ACE}(G;a_1,a_0)$;
6. Solve the optimization problem: $\min_G \max_{\lambda \geq 0} \max_D \{\mathrm{ACE}(G;a_1,a_0) + \lambda\,(V(G,D) - m - \eta)\}$;

Proof to Proposition 1

Proposition 1. *Let C be an instrumental variable for ACE of A on B, then both bounds computed from Problem 1 will converge to $\frac{\mathrm{cov}(B,C)}{\mathrm{cov}(A,C)}(a_1 - a_0)$ if the generated distribution converges to the observational distribution.*

Proof. Without loss of generality, we present our proof using the example in Figure 1 in the main paper, where C is an instrumental variable for ACE of A on B. The linear parameterized causal model is given by

$$c = g_C(\mathbf{z}_1), \quad a = \theta_A \cdot c + g_A(\mathbf{z}_2), \quad b = \theta_B \cdot a + g_B(\mathbf{z}_2)$$

The ACE can be directly obtained as $\mathbb{E}[\theta_B \cdot a_1 + g_B(\mathbf{z}_2)] - \mathbb{E}[\theta_B \cdot a_0 + g_B(\mathbf{z}_2)] = \theta_B \cdot (a_1 - a_0)$. As the generated distribution converges to the observational distribution, we have that $P(g_C(\mathbf{z}_1))$ with $\mathbf{z}_1$ following $P(\mathbf{z}_1)$ will converge to $P(c)$. Meanwhile, $P(\theta_A \cdot c + g_A(\mathbf{z}_2))$ and $P(\theta_B \cdot a + g_B(\mathbf{z}_2))$ with $\mathbf{z}_1$ following $P(\mathbf{z}_1)$ and $\mathbf{z}_2$ following $P(\mathbf{z}_2)$ will converge to $P(a)$ and $P(b)$ respectively, i.e.,

$$g_C(\mathbf{z}_1) \xrightarrow{\mathbf{z}_1 \sim P(\mathbf{z}_1)} P(c) \tag{1}$$

$$\theta_A g_C(\mathbf{z}_1) + g_A(\mathbf{z}_2) \xrightarrow{\mathbf{z}_1 \sim P(\mathbf{z}_1),\, \mathbf{z}_2 \sim P(\mathbf{z}_2)} P(a) \tag{2}$$

$$\theta_B(\theta_A g_C(\mathbf{z}_1) + g_A(\mathbf{z}_2)) + g_B(\mathbf{z}_2) \xrightarrow{\mathbf{z}_1 \sim P(\mathbf{z}_1),\, \mathbf{z}_2 \sim P(\mathbf{z}_2)} P(b) \tag{3}$$

It follows that

$$\mathbb{E}_{\mathbf{z}_1}[g_C(\mathbf{z}_1)] = \mathbb{E}[C] \tag{4}$$

$$\theta_A \mathbb{E}_{\mathbf{z}_1}[g_C(\mathbf{z}_1)] + \mathbb{E}_{\mathbf{z}_2}[g_A(\mathbf{z}_2)] = \mathbb{E}[A] \tag{5}$$

$$\theta_B(\theta_A \mathbb{E}_{\mathbf{z}_1}[g_C(\mathbf{z}_1)] + \mathbb{E}_{\mathbf{z}_2}[g_A(\mathbf{z}_2)]) + \mathbb{E}_{\mathbf{z}_2}[g_B(\mathbf{z}_2)] = \mathbb{E}[B] \tag{6}$$

In addition, from Eqs. (1) and (2) we have

$$\theta_A \mathbb{E}_{\mathbf{z}_1}[g_C^2(\mathbf{z}_1)] + \mathbb{E}_{\mathbf{z}_1,\mathbf{z}_2}[g_C(\mathbf{z}_1) g_A(\mathbf{z}_2)] = \mathbb{E}[AC] \tag{7}$$

and from Eqs. (1) and (3) we have

$$\begin{aligned} \theta_B(\theta_A \mathbb{E}_{\mathbf{z}_1}[g_C^2(\mathbf{z}_1)] + \mathbb{E}_{\mathbf{z}_1,\mathbf{z}_2}[g_C(\mathbf{z}_1) g_A(\mathbf{z}_2)]) \\ + \mathbb{E}_{\mathbf{z}_1,\mathbf{z}_2}[g_C(\mathbf{z}_1) g_B(\mathbf{z}_2)] = \mathbb{E}[BC]. \end{aligned} \tag{8}$$

By combining Eqs. (4), (5) and (7) we have $\theta_A \mathrm{var}(g_C(\mathbf{z}_1)) = \mathrm{cov}(A, C)$, and from Eqs. (4), (6) and (8) we have $\theta_B \theta_A \mathrm{var}(g_C(\mathbf{z}_1)) = \mathrm{cov}(B, C)$. As a result, we have that θ_B converges to $\frac{\mathrm{cov}(B,C)}{\mathrm{cov}(A,C)}$, meaning that the ACE converges to $\frac{\mathrm{cov}(B,C)}{\mathrm{cov}(A,C)}(a_1 - a_0)$ for both bounds. □

Experiments

All the experiments were conducted in an Ubuntu 18.04 PC with Intel Core i7-9700K and 16GB RAM. The implementation was developed using Ananconda, an open-source distribution of Python 3.8. The neural networks were implemented in PyTorch (Paszke et al. 2017).

Synthetic Data with Linear Settings

To evaluate the special linear case, we slightly modify the causal model of the synthetic data by substituting the non-linear W equation with $w = \theta_0^w + \theta_1^w x + \theta_2^w u_2 + \epsilon_w$. Similarly, we generate 10,000 samples with the other settings

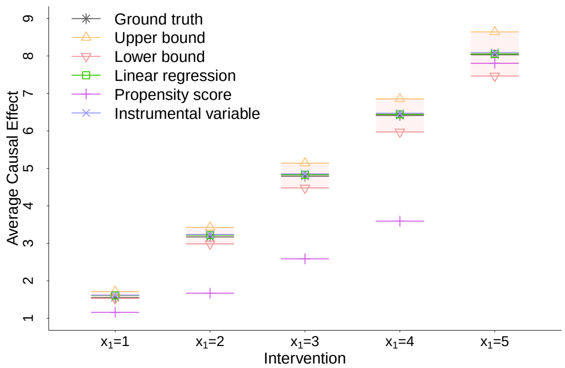

Figure 1: Average causal effects with different interventions ($x_0 = 0$) on the linear synthetic dataset.

unchanged. The results of experiments in the linear setting are shown in Figure 1. As can be seen, our upper bound and lower bound cover the ground truth in all intervention cases, and the range between the bounds is significantly reduced compared with the non-linear setting. The range is not negligible in some cases, probably due to the randomness in the training process and the imperfect fitting to the observational distribution. We expect that stronger generators may produce tighter bounds, and will evaluate different generators in our future work. As for baseline methods, the propensity score method still does not perform well, but the linear regression and instrumental variable methods can produce good estimations due to the satisfaction of the linear assumption.

References

Paszke, A.; Gross, S.; Chintala, S.; Chanan, G.; Yang, E.; DeVito, Z.; Lin, Z.; Desmaison, A.; Antiga, L.; and Lerer, A. 2017. Automatic differentiation in pytorch .