PAC Style Guarantees for Doubly Robust Generalized Front-Door Estimator

Yuhao Wang [1] Arnab Bhattacharyya [1] Jin Tian [2] N. V. Vinodchandran [3]

[1] National University of Singapore [2] Iowa State University

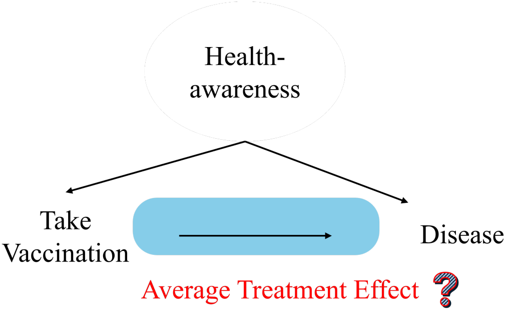
[3] University of Nebraska-Lincoln

Applications of Causal Inference

- **Healthcare:**
  - Understanding the effect of treatments and interventions.
  - Example: Determining if a new drug reduces the risk of a disease.
- **Public Policy:**
  - Evaluating the impact of policies and programs.
  - Example: Assessing the effectiveness of a new education policy on student performance.
- **Business:**
  - Identifying strategies that increase sales or customer retention.
  - Example: Measuring the effect of a marketing campaign on product sales.

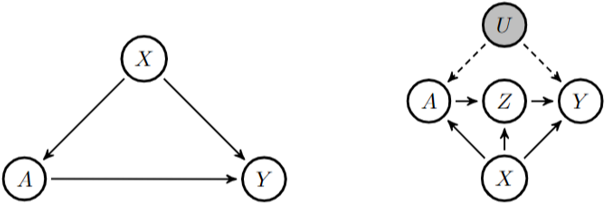
Background on Causal Inference

Example: understand average treatment effect in health care



Background: Plausible Causal Models

The causal effect can be identified using backdoor, front door models



(a) G_1: Back-door   (b) G_2: Front door

Background: ATE Estimation

Example: ATE estimation with back-door model

- **Outcome Regression Model**
  - Predicts the potential outcomes given covariates and treatment.
  - Estimates $E[Y|A, X]$.

Background: ATE Estimation

Example: ATE estimation with back-door model

- **Outcome Regression Model**
  - Predicts the potential outcomes given covariates and treatment.
  - Estimates $E[Y|A, X]$.
- **Propensity Score Model**
  - Estimates the probability of receiving the treatment given covariates.
  - Formally, $e(X) = P(A = 1|X)$.

Background: ATE Estimation

Example: ATE estimation with back-door model

- **Outcome Regression Model**
  - Predicts the potential outcomes given covariates and treatment.
  - Estimates $E[Y|A, X]$.
- **Propensity Score Model**
  - Estimates the probability of receiving the treatment given covariates.
  - Formally, $e(X) = P(A = 1|X)$.

However:

- Model misspecification is common $\Rightarrow$ Incorrect estimation.

Background: ATE Estimation

Example: ATE estimation with back-door model

- **Outcome Regression Model**
  - Predicts the potential outcomes given covariates and treatment.
  - Estimates $E[Y|A, X]$.
- **Propensity Score Model**
  - Estimates the probability of receiving the treatment given covariates.
  - Formally, $e(X) = P(A = 1|X)$.

However:

- Model misspecification is common $\Rightarrow$ Incorrect estimation.
- Slow convergence is common $\Rightarrow$ Not $\sqrt{n}$-consistency

Introduction to Doubly Robust Estimators

- **Recent advance: Doubly robust estimator**
  - Doubly robust estimators combine inverse probability weighting (IPW) and outcome regression to estimate causal effects;

Introduction to Doubly Robust Estimators

- **Recent advance: Doubly robust estimator**
  - Doubly robust estimators combine inverse probability weighting (IPW) and outcome regression to estimate causal effects;
  - Provides a general framework for estimating treatments effects using ML methods;

Introduction to Doubly Robust Estimators

- **Recent advance: Doubly robust estimator**
  - Doubly robust estimators combine inverse probability weighting (IPW) and outcome regression to estimate causal effects;
  - Provides a general framework for estimating treatments effects using ML methods;
    - Can use any (Preferably $n^{1/4}$-consistent) ML estimator with this approach

Introduction to Doubly Robust Estimators

- **Recent advance: Doubly robust estimator**
  - Doubly robust estimators combine inverse probability weighting (IPW) and outcome regression to estimate causal effects;
  - Provides a general framework for estimating treatments effects using ML methods;
    - Can use any (Preferably $n^{1/4}$-consistent) ML estimator with this approach
  - Enable us to construct valid confidence interval for our treatment effect estimates;

Introduction to Doubly Robust Estimators

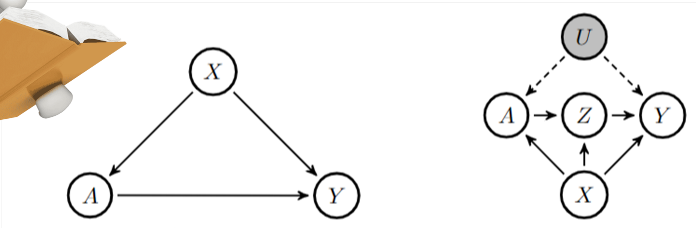
- **Recent advance: Doubly robust estimator**
  - Doubly robust estimators combine inverse probability weighting (IPW) and outcome regression to estimate causal effects;
  - Provides a general framework for estimating treatments effects using ML methods;
    - Can use any (Preferably $n^{1/4}$-consistent) ML estimator with this approach
  - Enable us to construct valid confidence interval for our treatment effect estimates;
  - Introduce a $\sqrt{n}$-consistent estimator

Introduction to Doubly Robust Estimators

- **Recent advance: Doubly robust estimator**
  - Doubly robust estimators combine inverse probability weighting (IPW) and outcome regression to estimate causal effects;
  - Provides a general framework for estimating treatments effects using ML methods;
    - Can use any (Preferably $n^{1/4}$-consistent) ML estimator with this approach
  - Enable us to construct valid confidence interval for our treatment effect estimates;
  - Introduce a $\sqrt{n}$-consistent estimator
    - As $n \to \infty$, the estimation error $\widehat{\psi} - \psi^*$ goes to zero at a rate of $n^{-1/2}$;
    - We really like our estimators to be at least $\sqrt{n}$-consistent.

Background: ATE Estimation

Apply doubly robust estimator!!



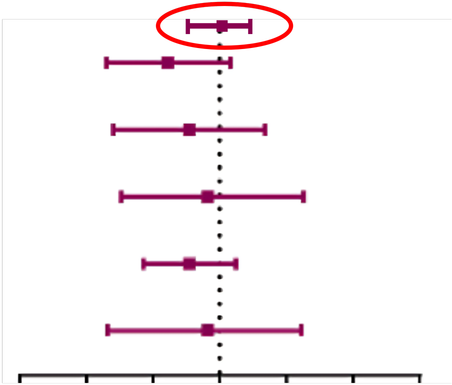
(a) G_1: Back-door (b) G_2: Front door

Our Motivation

However, only have finite observations

	Take vaccinatio n	Take placebo	Individual Treatment effect
	15.48	-	-
	13.52	-	-
	-	16.2	-
...	-	9.58	-
	...	...	...
Average	12.7	10.52	2.18 ?

Our Motivation



Background: Introduction to PAC Learning

Probably Approximately Correct (PAC) Learning is a framework provides a theoretical foundation for understanding finite sample complexity.

- Accuracy (ϵ): the maximum allowed error.
- Confidence (δ): the probability that the learned hypothesis is approximately correct.
- Sample complexity: the number of samples(n) required to achieve (ϵ, δ) guarantees.

Background: Introduction to PAC Learning

Probably Approximately Correct (PAC) Learning is a framework provides a theoretical foundation for understanding finite sample complexity.

- Accuracy (ϵ): the maximum allowed error.
- Confidence (δ): the probability that the learned hypothesis is approximately correct.
- Sample complexity: the number of samples(n) required to achieve (ϵ, δ) guarantees.

Goal: With high probability ("Probably"), the selected hypothesis will have lower error ("Approximately Correct")

Our Problem

- PAC-style guarantees for doubly machine learning estimator

Back-door Model

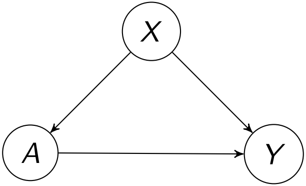


Figure: Back-door adjustment

- $\mathcal{A}$ be $\{0, 1\}$
- $\mathcal{X}, \mathcal{Y}$ be any sets
- X be a random variable on $\mathcal{X}$
- $A(x), \widehat{A}(x)$ be random variables on $\mathcal{A}$ for any $x \in \mathcal{X}$
- $Y(a, x), \widehat{Y}(a, x)$ be random variables on $\mathcal{Y}$ for any $a \in \mathcal{A}, x \in \mathcal{X}$

Back-door adjustment: Estimator

The causal effect of A on Y denote as:

$$\psi^* = \mathbb{E}[Y \mid \text{do}(A = a^*)]$$

we define:

$$\phi(a, x, y) = \mathop{\mathbb{E}}_{\widehat{Y}(a^*, x)} \widehat{Y} + \frac{1[a = a^*]}{\widehat{\mathcal{A}}(a; x)} \left(y - \mathop{\mathbb{E}}_{\widehat{Y}(a, x)} \widehat{Y} \right)$$

and

$$\psi_n = \frac{1}{n} \sum_{i=1}^{n} \phi_{\widehat{A}, \widehat{Y}}(a_i, x_i, y_i)$$

Back-door adjustment: Goal

We revisit the well-known double/debiased machine learning (DML) estimator for covariate adjustment in the BD setting [RRZ94; Che+17] and analyze the mean-squared error of $|\psi_n - \psi^*|$ in the finite sample setting.

Specifically, given finite samples, we want to bound MSE error in term of the error in $\widehat{A}$, $\widehat{Y}$.

Back-door adjustment: Result

Definition

Given two distributions P and Q, and an even integer $p \geq 2$, we define the χ^p divergence between them as:

$$\chi^p(P \| Q) = \mathbb{E}_{x \sim P}\left[\left(1 - \frac{Q(x)}{P(x)}\right)^p\right].$$

Novelty: The novelty of our result is that we express the mean-squared error explicitly in terms of the errors in the estimates of the treatment and outcome distributions. These errors are formulated in terms of χ^2-divergence.

Back-door adjustment: Result

Assumption

Assume for all x, the following condition holds:

$$\mathop{\mathbb{E}}_{Y(a^*,x)} Y^2 \le V, \quad \Pr[A(x) = a^*] \ge \mu, \text{ and } \quad \Pr[\widehat{A}(x) = a^*] \ge \mu.$$

Theorem

Under the above Assumption, for any $\varepsilon > 0$:

$$\Pr[|\psi_n - \psi^*| > \varepsilon] < \frac{1}{n\varepsilon^2}\mathcal{O}_{V,\mu}\left(1 + \mathop{\mathbb{E}}_{x}\chi^2\left(\widehat{Y}(a^*,x)\|Y(a^*,x)\right) + \mathop{\mathbb{E}}_{x}\chi^2\left(\widehat{A}(x)||A(x)\right)\right)$$
$$+ \frac{1}{\varepsilon^2}\mathcal{O}_{V,\mu}\left(\mathop{\mathbb{E}}_{x}\chi^2\left(\widehat{Y}(a^*,x)\|Y(a^*,x)\right)\cdot\chi^2\left(\widehat{A}(x)\|A(x)\right)\right)$$

Back-door adjustment: Result

What does this result mean?

$$\Pr[|\psi_n - \psi^*| > \varepsilon] < \frac{1}{n\varepsilon^2}\mathcal{O}_{V,\mu}\left(\underbrace{1}_{Part1} + \underbrace{\mathop{\mathbb{E}}_{x}\chi^2\left(\widehat{Y}(a^*,x)\|Y(a^*,x)\right)}_{Part2} + \underbrace{\mathop{\mathbb{E}}_{x}\chi^2\left(\widehat{A}(x)\|A(x)\right)}_{Part3}\right) + \underbrace{\frac{1}{\varepsilon^2}\mathcal{O}_{V,\mu}\left(\mathop{\mathbb{E}}_{x}\chi^2\left(\widehat{Y}(a^*,x)\|Y(a^*,x)\right)\cdot\chi^2\left(\widehat{A}(x)\|A(x)\right)\right)}_{Part4}$$

Back-door adjustment: Result

What does this result mean?

$$
\begin{aligned}
&\Pr[|\psi_n - \psi^*| > \varepsilon] \\
&< \frac{1}{n\varepsilon^2}\mathcal{O}_{V,\mu}\left(\underbrace{1}_{Part1} + \underbrace{\mathop{\mathbb{E}}_{x}\chi^2\left(\widehat{Y}(a^*,x)\|Y(a^*,x)\right)}_{Part2} + \underbrace{\mathop{\mathbb{E}}_{x}\chi^2\left(\widehat{A}(x)\|A(x)\right)}_{Part3}\right) \\
&\quad + \underbrace{\frac{1}{\varepsilon^2}\mathcal{O}_{V,\mu}\left(\mathop{\mathbb{E}}_{x}\chi^2\left(\widehat{Y}(a^*,x)\|Y(a^*,x)\right)\cdot\chi^2\left(\widehat{A}(x)\|A(x)\right)\right)}_{Part4}
\end{aligned}
$$

- Part 1: error incurred by the oracle estimator;

Back-door adjustment: Result

What does this result mean?

$$\Pr[|\psi_n - \psi^*| > \varepsilon] < \frac{1}{n\varepsilon^2}\mathcal{O}_{V,\mu}\left(\underbrace{1}_{Part1} + \underbrace{\mathop{\mathbb{E}}_{x}\chi^2\left(\widehat{Y}(a^*,x)\|Y(a^*,x)\right)}_{Part2} + \underbrace{\mathop{\mathbb{E}}_{x}\chi^2\left(\widehat{A}(x)\|A(x)\right)}_{Part3}\right) + \underbrace{\frac{1}{\varepsilon^2}\mathcal{O}_{V,\mu}\left(\mathop{\mathbb{E}}_{x}\chi^2\left(\widehat{Y}(a^*,x)\|Y(a^*,x)\right)\cdot\chi^2\left(\widehat{A}(x)\|A(x)\right)\right)}_{Part4}$$

- Part 1: error incurred by the oracle estimator;
- Part 2 & Part 3: Mismatch measured in χ^2 between the model estimates and the truth, for the outcome and propensity distributions;

Back-door adjustment: Result

What does this result mean?

$$\Pr[|\psi_n - \psi^*| > \varepsilon] < \frac{1}{n\varepsilon^2}\mathcal{O}_{V,\mu}\left(\underbrace{1}_{Part1} + \underbrace{\mathop{\mathbb{E}}_{x}\chi^2\left(\widehat{Y}(a^*,x)\|Y(a^*,x)\right)}_{Part2} + \underbrace{\mathop{\mathbb{E}}_{x}\chi^2\left(\widehat{A}(x)\|A(x)\right)}_{Part3}\right) + \underbrace{\frac{1}{\varepsilon^2}\mathcal{O}_{V,\mu}\left(\mathop{\mathbb{E}}_{x}\chi^2\left(\widehat{Y}(a^*,x)\|Y(a^*,x)\right)\cdot\chi^2\left(\widehat{A}(x)\|A(x)\right)\right)}_{Part4}$$

- Part 1: error incurred by the oracle estimator;
- Part 2 & Part 3: Mismatch measured in χ^2 between the model estimates and the truth, for the outcome and propensity distributions;
- Part 4: *Mixed-bias* or *product rate* phenomenon of the doubly robust estimators [Che+20]

Front-door Criterion - Model

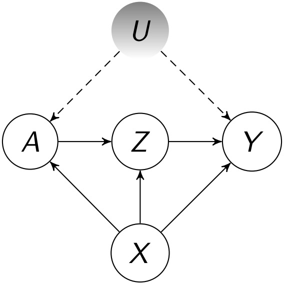


Figure: Front-door adjustment

- $\mathcal{A}$ be $\{0, 1\}$, and $\mathcal{X}, \mathcal{Z}, \mathcal{Y}$ be any sets
- X be a random variable on $\mathcal{X}$
- $A(x), \widehat{A}(x)$ be random variables on $\mathcal{A}$ for any $x \in \mathcal{X}$
- $Z(a, x), \widehat{Z}(z, x)$ be random variable on $\mathcal{Z}$ for any $a \in \mathcal{A}, x \in \mathcal{X}$
- $Y(z, a, x), \widehat{Y}(z, a, x)$ be random variables on $\mathcal{Y}$ for any $a \in \mathcal{A}, z \in \mathcal{Z}, x \in \mathcal{X}$

Front-door adjustment - Notations

- we use ψ^* denote the ground truth estimator
- $\phi_{Y,Z,A}$ to denote the oracle estimator
- $\phi_{\widehat{Y},\widehat{Z},\widehat{A}}$ denote the empirical estimator.
- $\psi^* = \mathbb{E}[Y \mid \mathrm{do}(A = a^*)]$
- $\psi_n = \frac{1}{n}\sum_{i=1}^{n} \phi(a_i, z_i, y_i, x_i)$

Our goal is to bound the MSE error of $(\psi_n - \psi^*)$ in terms of the error in $\widehat{A}, \widehat{Z}, \widehat{Y}$.

Front-door adjustment- Estimators

ψ^* can be identified in terms of the observable distribution in a number of ways, all equivalent to each other.

- The first, ψ_1, doesn't require $A \mid X$:

$$\psi_1 = \mathop{\mathbb{E}}_{A,X}\left[\mathop{\mathbb{E}}_{Z}\left[\mathop{\mathbb{E}}_{Y}[Y \mid A, Z, X] \mid a^*, X\right]\right].$$

- The second, ψ_2, doesn't require $Z \mid A, X$:

$$\psi_2 = \mathop{\mathbb{E}}_{A,Z,X}\left[\frac{1[A = a^*]}{\Pr[A \mid X]} \cdot \mathop{\mathbb{E}}_{A'}\left[\mathop{\mathbb{E}}_{Y}[Y \mid A', Z, X] \mid X\right]\right].$$

- The third, ψ_3, doesn't require $Y \mid A, Z, X$:

$$\psi_3 = \mathop{\mathbb{E}}_{Y,A,Z,X}\left[Y \cdot \frac{\Pr[Z \mid a^*, X]}{\Pr[Z \mid A, X]}\right]$$

Front-door adjustment - DR estimator

Define the following quantity [Ful+20]:

$$\begin{aligned}
&\phi(a, z, y, x) \\
&= \mathop{\mathbb{E}}_{\widehat{Z}(a^*,x)} \left[\mathop{\mathbb{E}}_{\widehat{Y}(a,\widehat{Z},x)} \widehat{Y} \right] + \frac{1[a = a^*]}{\widehat{\mathcal{A}}(a; x)} \mathop{\mathbb{E}}_{\widehat{A}(x)} \left(\mathop{\mathbb{E}}_{\widehat{Y}(\widehat{A},z,x)} \widehat{Y} - \mathop{\mathbb{E}}_{\widehat{Z}(a,x)} \left[\mathop{\mathbb{E}}_{\widehat{Y}(\widehat{A},\widehat{Z},x)} \widehat{Y} \right] \right) \\
&\qquad + \left(y - \mathop{\mathbb{E}}_{\widehat{Y}(a,z,x)} \widehat{Y} \right) \cdot \frac{\widehat{\mathcal{Z}}(z; a^*, x)}{\widehat{\mathcal{Z}}(z; a, x)}
\end{aligned}$$

We can get a "doubly-robustness" type property for ϕ: If any two of $\widehat{A}, \widehat{Z}, \widehat{Y}$ are always correct, then $\mathbb{E}_{A,Z,Y,X}\, \phi = \psi^*$.

Front-door adjustment - Our Result

Assumption

Assume $\forall a, z, x,$

$$\mathop{\mathbb{E}}_{Y(a,z,x)} Y^2 \leq V, \;\; \Pr[Z(a,x) = z] \geq \mu_Z \cdot 1[\Pr[Z(a^*,x) = z] \neq 0],$$
$$\textit{and} \quad \Pr[A(x) = a^*] \geq \mu_A$$

Front-door adjustment - Our results

Theorem

$$
\begin{aligned}
&\Pr[|\psi_n - \psi^*| > \varepsilon] \\
&< \frac{1}{n\varepsilon^2}\mathcal{O}_{V,\mu_Z,\mu_A}\Bigg(1 + \mathop{\mathbb{E}}_{a,z,x}\chi^2\left(\widehat{Y}(a,z,x)\|Y(a,z,x)\right) + \mathop{\mathbb{E}}_{x}\chi^2\left(\widehat{A}(x)\|A(x)\right) \\
&+ \mathop{\mathbb{E}}_{x}\chi^2\left(\widehat{Z}(a^*,x)\|Z(a^*,x)\right) + \mathop{\mathbb{E}}_{a,x}\chi^2\left(\widehat{Z}(a,x)\|Z(a,x)\right)\Bigg) \\
&+ \frac{1}{\varepsilon^2}\mathcal{O}_{V,\mu_A,\mu_Z}\Bigg(\Bigg(\sqrt{\mathbb{E}_{a,z,x}\chi^4(\widehat{Y}(a,z,x)\|Y(a,z,x))} + \sqrt{\mathbb{E}_x\chi^4(\widehat{A}(x)\|A(x))} \\
&\quad + \sqrt{\mathbb{E}_x\chi^4(\widehat{Z}(a^*,x)\|Z(a^*,x))} + \sqrt{\mathbb{E}_{a,x}\chi^4(\widehat{Z}(a,x)\|Z(a,x))}\Bigg)^2\Bigg)
\end{aligned}
$$

Front-door adjustment - Our Result

What does this result mean?

$$\Pr[|\psi_n - \psi^*| > \varepsilon]$$

$$< \frac{1}{n\varepsilon^2}\mathcal{O}_{V,\mu_Z,\mu_A}\Bigg(\underbrace{1}_{Oracle} + \underbrace{\mathop{\mathbb{E}}_{a,z,x}\chi^2\left(\widehat{Y}(a,z,x)\|Y(a,z,x)\right)}_{\text{Outcome mismatch}} + \underbrace{\mathop{\mathbb{E}}_{x}\chi^2\left(\widehat{A}(x)\|A(x)\right)}_{\text{Propensity mismatch}}$$

$$+ \underbrace{\mathop{\mathbb{E}}_{x}\chi^2\left(\widehat{Z}(a^*,x)\|Z(a^*,x)\right)}_{\text{Propensity mismatch}} + \underbrace{\mathop{\mathbb{E}}_{a,x}\chi^2\left(\widehat{Z}(a,x)\|Z(a,x)\right)}_{\text{Propensity mismatch}}\Bigg)$$

$$+ \frac{1}{\varepsilon^2}\mathcal{O}_{V,\mu_A,\mu_Z}\Bigg(\Bigg(\sqrt{\mathbb{E}_{a,z,x}\chi^4(\widehat{Y}(a,z,x)\|Y(a,z,x))} + \sqrt{\mathbb{E}_x\chi^4(\widehat{A}(x)\|A(x))}$$

$$+ \sqrt{\mathbb{E}_x\chi^4(\widehat{Z}(a^*,x)\|Z(a^*,x))} + \sqrt{\mathbb{E}_{a,x}\chi^4(\widehat{Z}(a,x)\|Z(a,x))}\Bigg)^2\Bigg)$$

Conclusion

Summary

- Our bound can be used to get guidance on how to construct the estimators $\widehat{Y}$, $\widehat{Z}$, and $\widehat{A}$;

Conclusion

Summary

- Our bound can be used to get guidance on how to construct the estimators $\widehat{Y}$, $\widehat{Z}$, and $\widehat{A}$;
- Learning a distribution that minimizes a particular divergence is a question in *distribution learning*

Conclusion

Summary

- Our bound can be used to get guidance on how to construct the estimators $\widehat{Y}$, $\widehat{Z}$, and $\widehat{A}$;
- Learning a distribution that minimizes a particular divergence is a question in *distribution learning*
- For example, the problem of learning a distribution minimizing the χ^2 divergence was explicitly studied in [Kam+15].

Future Work

- Extend the current work to more general graph?

Future Work

- Extend the current work to more general graph?
- How to analyze the mixed-bias term in a "simpler way"?

Future Work

- Extend the current work to more general graph?
- How to analyze the mixed-bias term in a "simpler way"?
- Real-data experiments.

Thank You!

Q & A

References I

[Che+17] Victor Chernozhukov et al. “Double/debiased/neyman machine learning of treatment effects”. In: *American Economic Review* 107.5 (2017), pp. 261–65.

[Che+20] Victor Chernozhukov et al. “Adversarial estimation of riesz representers”. In: *arXiv preprint arXiv:2101.00009* (2020).

[Ful+20] Isabel R Fulcher et al. “Robust inference on population indirect causal effects: the generalized front door criterion”. In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 82.1 (2020), pp. 199–214.

[Kam+15] Sudeep Kamath et al. “On learning distributions from their samples”. In: *Conference on Learning Theory*. PMLR. 2015, pp. 1066–1100.

References II

[RRZ94] James M Robins, Andrea Rotnitzky, and Lue Ping Zhao. "Estimation of regression coefficients when some regressors are not always observed". In: *Journal of the American statistical Association* 89.427 (1994), pp. 846–866.

Backup

Back-door adjustment

$$\Delta_1 = \frac{1}{n}\sum_i \left(\underset{\widehat{Y}(a^*,x_i)}{\mathbb{E}} \widehat{Y} - \underset{Y(a^*,x_i)}{\mathbb{E}} Y - \frac{1[a_i = a^*]}{\mathcal{A}(a_i;x_i)} \left(\underset{\widehat{Y}(a_i,x_i)}{\mathbb{E}} \widehat{Y} - \underset{Y(a_i,x_i)}{\mathbb{E}} Y \right) \right)$$

$$\Delta_2 = \frac{1}{n}\sum_i 1[a_i = a^*] \left(\frac{1}{\widehat{\mathcal{A}}(a_i;x_i)} - \frac{1}{\mathcal{A}(a_i;x_i)} \right) \cdot \left(y_i - \underset{Y(a_i,x_i)}{\mathbb{E}} Y \right)$$

$$\Delta_3 = -\frac{1}{n}\sum_i \left(\underset{\widehat{Y}(a_i,x_i)}{\mathbb{E}} \widehat{Y} - \underset{Y(a_i,x_i)}{\mathbb{E}} Y \right) \left(\frac{1}{\widehat{\mathcal{A}}(a_i;x_i)} - \frac{1}{\mathcal{A}(a_i;x_i)} \right)$$

Back-door adjustment

We bound each of $\Delta_1, \Delta_2, \Delta_3$ by explicit computation.

$$\mathbb{E}[\Delta_1^2] \leq \mathcal{O}_{V,\mu}\left(\frac{1}{n} \cdot \mathop{\mathbb{E}}_{x} \chi^2\left(\widehat{Y}(a^*, x) \| Y(a^*, x)\right)\right)$$

$$\mathbb{E}[\Delta_2^2] \leq \mathcal{O}_{V,\mu}\left(\frac{1}{n} \cdot \mathop{\mathbb{E}}_{x} \chi^2\left(\widehat{A}(x) || A(x)\right)\right)$$

$$\mathbb{E}[\Delta_3^2] \leq \mathcal{O}_{V,\mu}\left(\mathop{\mathbb{E}}_{x} \chi^2\left(\widehat{Y}(a^*, x) \| Y(a^*, x)\right) \cdot \mathop{\mathbb{E}}_{x} \chi^2\left(\widehat{A}(x) \| A(x)\right)\right)$$

Front-door adjustment

Δ_4 can be written explicitly as follows:

$$
\begin{aligned}
\Delta_4 = & \underbrace{\sum_{z'} v(z'; a^*, x) \sum_{y'} y' \cdot u(y'; a, z', x)}_{\Delta_{41}} \\
& \underbrace{- \frac{1[a = a^*]}{\mathcal{A}(a; x)} \sum_{a'} \mathcal{A}(a'; x) \Bigg(\sum_{z'} v(z'; a, x) \cdot \sum_{y'} y' \cdot u(y'; a', z', x) \Bigg)}_{\Delta_{42}} \\
& \underbrace{+ \frac{1[a = a^*]}{\mathcal{A}(a; x)} \sum_{a'} w(a'; x) \Bigg(\sum_{y'} y' \cdot u(y'; a', z, x)}_{\Delta_{43}} - \underbrace{\sum_{z'} \mathcal{Z}(z'; a, x) \sum_{y'} y' \cdot u(y'; a', z', x)}_{\Delta_{44}} \\
& \qquad \qquad - \underbrace{\sum_{z'} v(z'; a, x) \cdot \sum_{y'} y' \cdot \mathcal{Y}(y'; a', z', x) \Bigg)}_{\Delta_{45}} \\
& \underbrace{+ 1[a = a^*] \frac{w(a; x)}{\mathcal{A}^2(a; x)} \cdot \sum_{a'} \mathcal{A}(a', x) \Bigg(\sum_{z'} \mathcal{Z}(z'; a, x) \cdot \sum_{y'} y' \cdot u(y'; a', z', x)}_{\Delta_{46}} \\
& \qquad \qquad + \underbrace{\sum_{z'} v(z'; a, x) \cdot \sum_{y'} y' \cdot \mathcal{Y}(y'; a', z', x)}_{\Delta_{47}} - \underbrace{\sum_{y'} y' \cdot u(y'; a', z, x) \Bigg)}_{\Delta_{48}} \\
& \underbrace{- 1[a = a^*] \frac{w(a; x)}{\mathcal{A}^2(a; x)} \sum_{a'} w(a'; x) \Bigg(\sum_{y'} y' \cdot (\mathcal{Y} + u)(y'; a', z, x)}_{\Delta_{49}} \\
& \qquad \qquad - \underbrace{\sum_{z'} (\mathcal{Z} + v)(z'; a, x) \cdot \sum_{y'} y' \cdot (\mathcal{Y} + u)(y'; a', z', x) \Bigg)}_{\Delta_{50}} \\
& + \Big(v(z; a^*, x) \cdot \mathcal{Z}(z; a, x) - v(z; a, x) \cdot \mathcal{Z}(z; a^*, x) \Big) \\
& \quad \underbrace{\cdot \Bigg[- \Bigg(y - \sum_{y'} y' \cdot \mathcal{Y}(y'; a, z, x) \Bigg) \frac{v(z; a, x)}{\mathcal{Z}^3(z; a, x)}}_{\Delta_{51}} - \underbrace{\sum_{y'} y' \cdot u(y'; a, z, x) \frac{1}{(\mathcal{Z}^2 + v\mathcal{Z})(z; a, x)} \Bigg]}_{\Delta_{52}}
\end{aligned}
$$