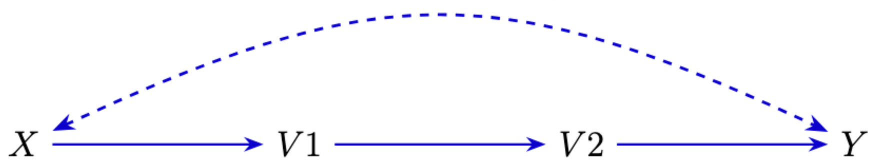
Variance Estimation and Selecting Estimands for Causal Effect Queries

Anna K Raichev (araichev@uci.edu), Rina Dechter (dechter@ics.uci.edu),
Jin Tian (jin.tian@mbzuai.ac.ae), and Alexander Ihler (ihler@ics.uci.edu)

The Problem

Given a causal diagram, an identifiable query $P(\boldsymbol{Y} \mid do(\boldsymbol{X}=x^*))$ and samples from the observed distribution, the task is to output the distribution of $E[\boldsymbol{Y} \mid do(\boldsymbol{X}=x^*)]$.

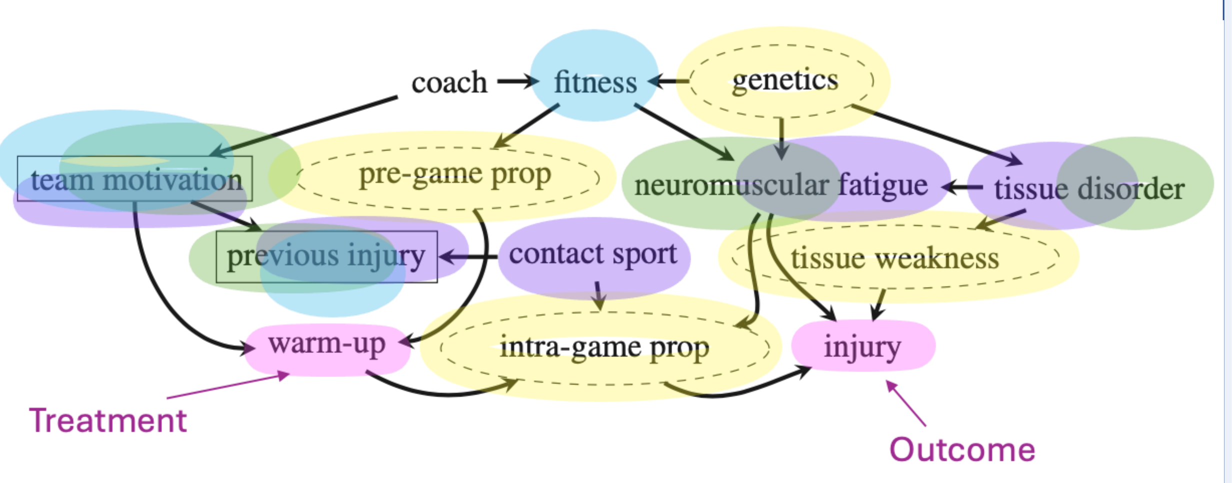


$$\Psi_{fd} = \mathbb{E}[Y \mid do(X=x^*)]$$
$$= \sum_{V1} P(V1|x^*) \sum_{X} \mathbb{E}[Y \mid X, V1] P(X)$$
$$= \sum_{V2} P(V2|x^*) \sum_{X} E[Y \mid X, V2] P(X)$$

Multiple estimands; which one should we chose ?

This work: smallest asymptotic variance

Motivating Example



Some adjustment sets:

Z_1 = {team motivation, previous injury, contact sport, tissue disorder, neuromuscular fatigue}

Z_2= {team motivation, previous injury, tissue disorder, neuromuscular fatigue}

Z_3 = {team motivation, previous injury, fitness}

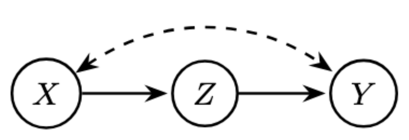
Which adjustment set is best?

Background

Structural Causal Model: $\boldsymbol{M} = \langle \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{F}, P(\boldsymbol{U}) \rangle$

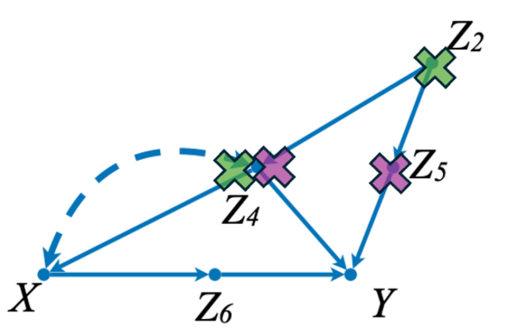
- $\boldsymbol{U} = \{U_1, \ldots, U_k\}$ set of unmeasurable latent variables
- $\boldsymbol{V} = \{V_1, V_2, \ldots, V_n\}$ set of observable variables
- $\boldsymbol{F} = \{f_i : V_i \in \boldsymbol{V}\}$ is a set of functional mechanisms f_i that each determine the value v_i of their corresponding V_i as a function of V_i's causal parents $PA_i \subseteq \boldsymbol{U} \cup \boldsymbol{V} \setminus V_i$
- $P(\boldsymbol{U})$ is a probability distribution over the exogenous variables

Causal Diagram:



Back-door Adjustments:

$$\Psi_{bd} = \sum_z \boldsymbol{E}[Y \mid x^*, z] P(z)$$

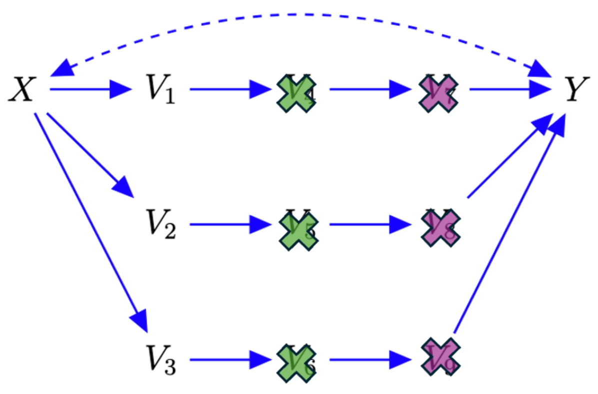


$Z = \{Z_2, Z_4\}$

$Z = \{Z_4, Z_5\}$

Front-door Adjustments:

$$\Psi_{fd} = \sum_z P(z \mid x^*) \sum_x \boldsymbol{E}[Y \mid x, z] P(x)$$



$Z = \{V_4, V_5, V_6\}$

$Z = \{V_7, V_8, V_9\}$

We say that adjustment set Z **dominates** Z' and write $Z \preccurlyeq Z'$ if:

- The asymptotic variance of the estimand using Z is always smaller or equal to Z' (using Z is equal or more efficient for all possible probability distributions)

Lemma 1 [Smucler et al 2022]

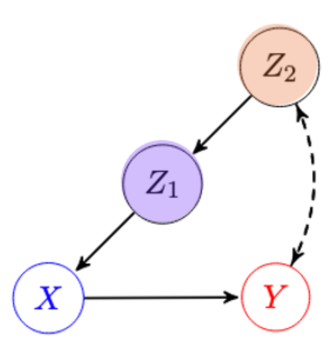
- Start with back-door set **B**
- Identify variables **A** independent of the treatment, X, given **B:**
  - X ⊥ ***A*** | ***B***
- A ∪ B is also a valid back-door adjustment set

$A = \{Z_2\}$

$B = \{Z_1\}$

A ∪ ***B*** ≼ ***B***

A ∪ ***B* dominates *B***



Lemma 2 [Smucler et al 2022]

- Start with back-door set ***A*** ∪ ***B*** with A and B disjoint
- Identify variables **B** independent of the outcome, Y, given treatment X and **A:**
  - Y⊥ ***B*** | ***X, A***
- **A** is also a valid back-door adjustment set

A ≼ ***A*** ∪ ***B***

A* dominates *A ∪ ***B***

Our Contribution

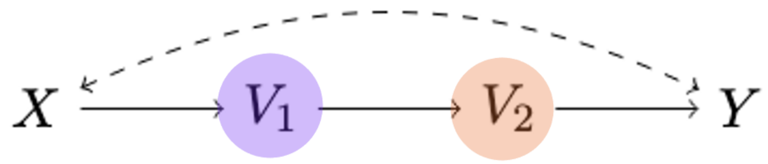
Theorem 1 *Given causal diagram $\mathcal{G}$ and identifiable query $P(Y \mid do(X))$, let $\boldsymbol{A} \cup \boldsymbol{B}$ be a* **front-door** *adjustment set with $\boldsymbol{A}$ and $\boldsymbol{B}$ disjoint, and $Y \perp\!\!\!\perp \boldsymbol{B} \mid \boldsymbol{A}, X$. Then $\boldsymbol{A}$ is also a front-door adjustment set and for any intervention $X = x^*$ and for all $P \in \mathcal{M}(\mathcal{G})$,*

$$\sigma^2_{x^*,fd}(\boldsymbol{A} \cup \boldsymbol{B}; P) - \sigma^2_{x^*,fd}(\boldsymbol{A}; P)$$
$$= \mathbb{E}\Big[\operatorname{Var}(Y|\, X, \boldsymbol{A}) \operatorname{Var}\Big[\frac{f(\boldsymbol{A}, \boldsymbol{B}|x^*)}{f(\boldsymbol{A}, \boldsymbol{B}|X)} \mid X, \boldsymbol{A}\Big]\Big] \geq 0.$$

Consequently, $\boldsymbol{A} \preccurlyeq \boldsymbol{A} \cup \boldsymbol{B}$.

$A = \{V_2\}$

$B = \{V_1\}$



- Start with front-door adjustment set ***A*** ∪ ***B***
- Identify variables **B** independent of the outcome, Y, given treatment X and **A:**
  - Y⊥ ***B*** | ***X, A***
- **A** is also a valid front-door adjustment set

A ≼ ***A*** ∪ ***B***

A* dominates *A ∪ ***B***

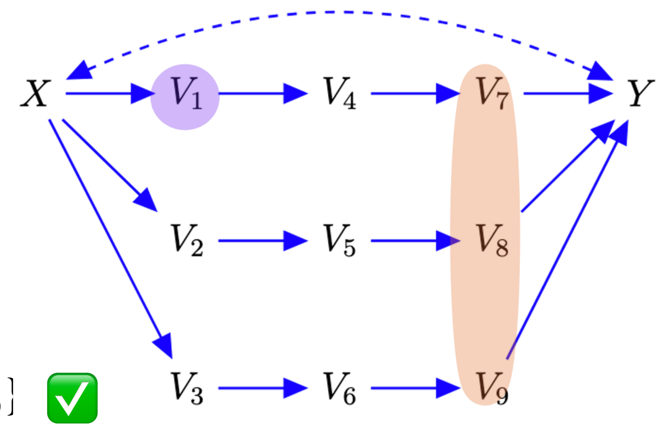
Examples

$A = \{V_7, V_8, V_9\}$

$B = \{V_1\}$

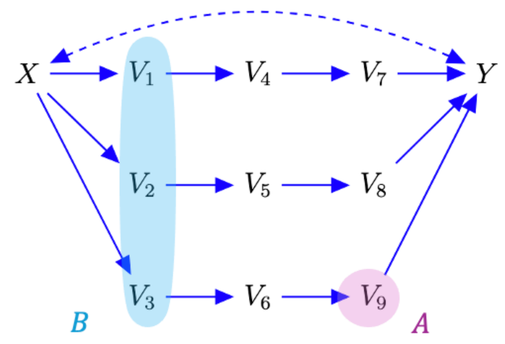
Should we add B to A?

1. Test: $Y \perp\!\!\!\perp B \mid A, X$
2. $Y \perp\!\!\!\perp \{V_1\} \mid \{X, V_7, V_8, V_9\}$ ✅



No, A **dominates** $A \cup B$

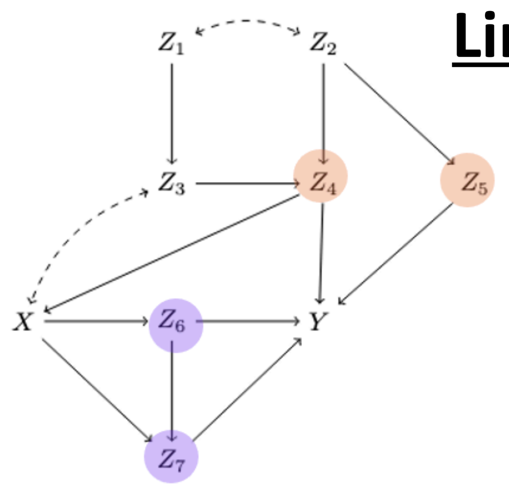
But, lemma 1 does NOT hold for front-doors



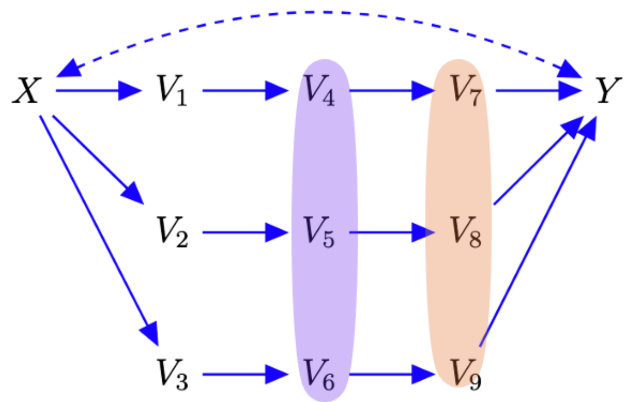
For some distributions **A** ∪ **B** has smaller variance,

And others **B** has smaller variance

Limitations



Front-door vs back-door sets



Disjoint sets

Known structural conditions don't apply

Can we estimate the variance to pick estimands?

Algorithms

Bootstrap Estimator for Variance:

Estimate the variance by **resampling the observed data**.

1. Construct the empirical distribution
2. Draw B bootstrap datasets
3. Compute the estimand on each sample
4. Compute empirical variance

Advantages:

- No distributional assumptions
- Works for complex estimands

Disadvantages:

- Computationally intensive

Plug-In Estimator for Variance:

Use **fisher information matrices** to estimate variance analytically using the empirical distribution [Fisher, 1922]

1. Use the MLE estimates on m samples
2. Calculate the fisher information matrix (I)
3. Compute the derivative:

$$\operatorname{Var}[\hat{\psi}(\rho)] = \nabla_\rho \hat{\psi}(\rho)^\top I(\rho)^{-1} \nabla_\rho \hat{\psi}(\rho),$$

Fisher Info Matrices

Computed with autograd

Advantages:

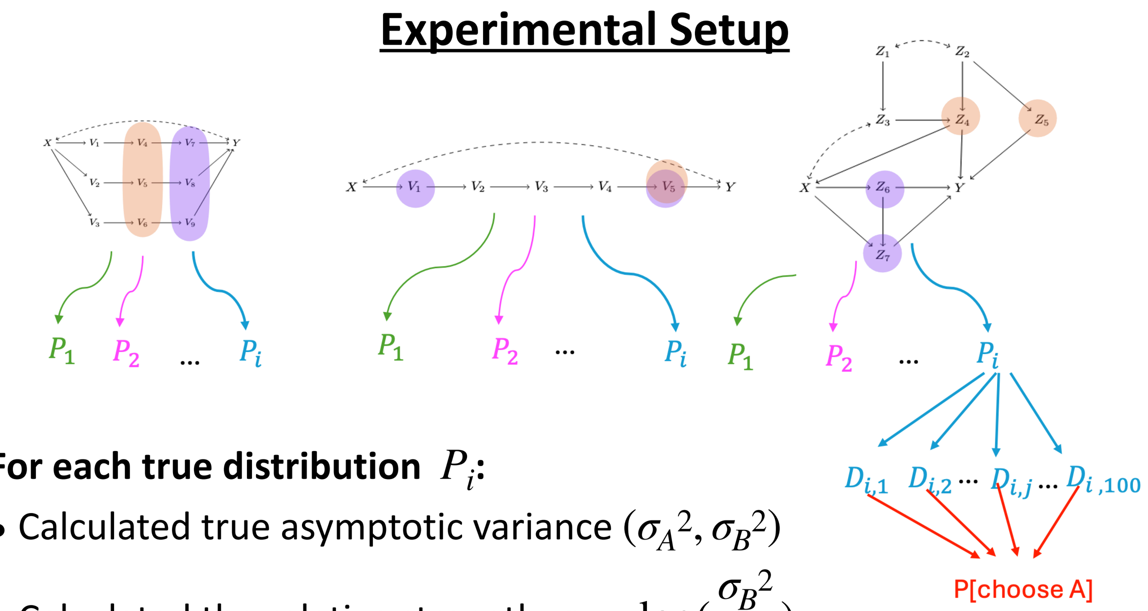
- Computationally efficient

Disadvantages:

- Not unbiased for finite samples

Empirical Evaluation

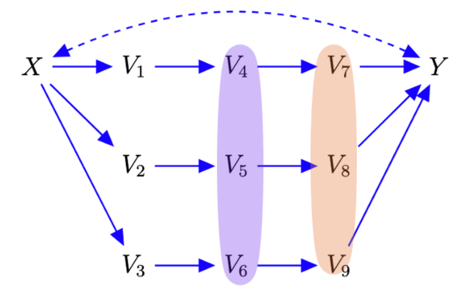
Experimental Setup



For each true distribution P_i:

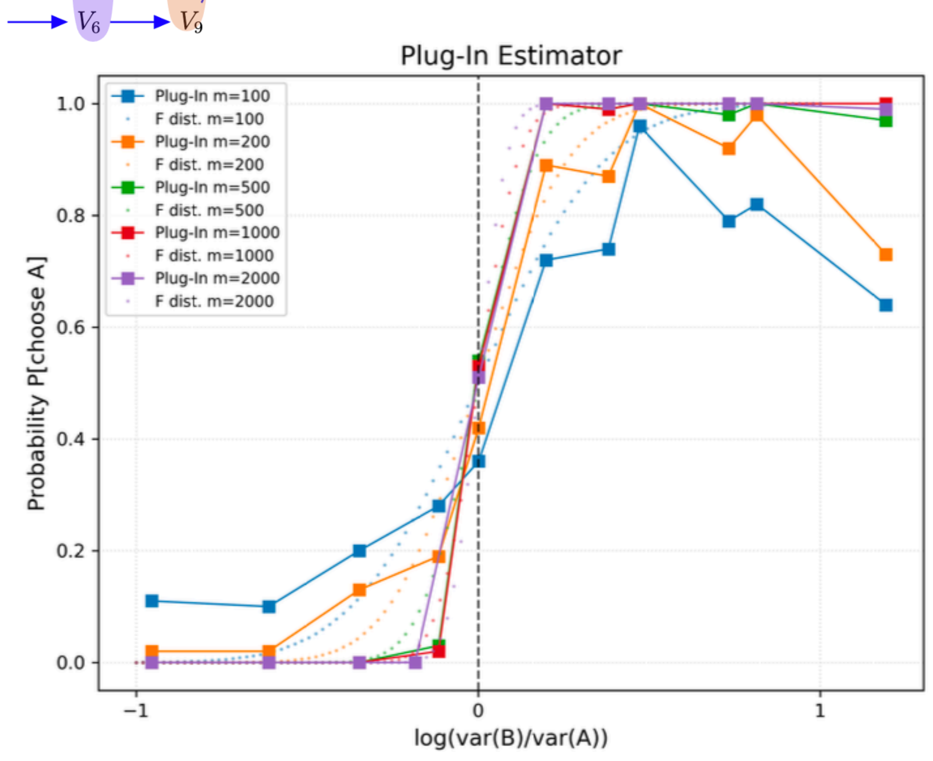
- Calculated true asymptotic variance $(\sigma_A{}^2, \sigma_B{}^2)$
- Calculated the relative strength: $s_i = \log(\frac{\sigma_B{}^2}{\sigma_A{}^2})$
- Sampled a collection of data sets $\{D_{ij}\}$ each of size m
- Calculate probability of choosing A over B for each estimator on $\{D_{ij}\}$

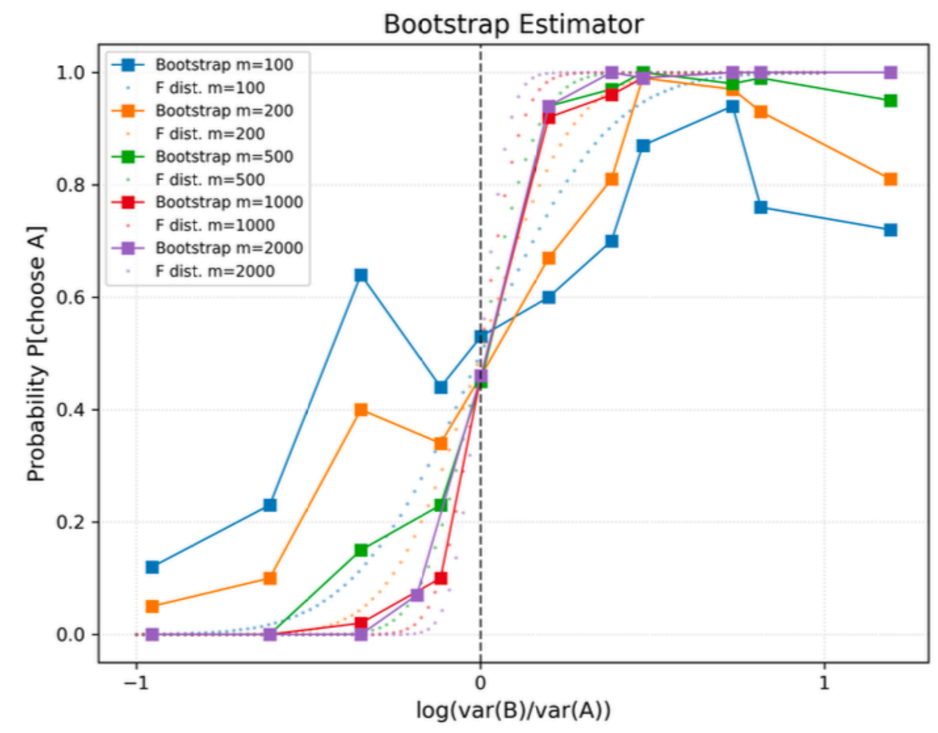
Results

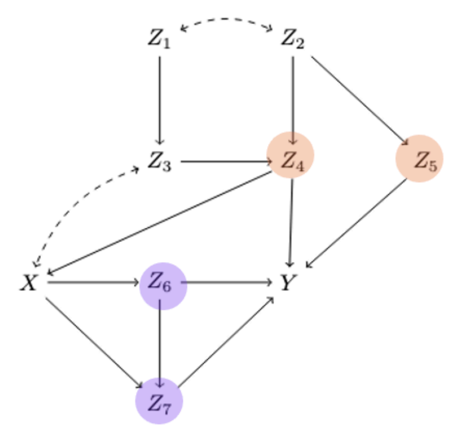


$A = \{V_7, V_8, V_9\}$

$B = \{V_4, V_5, V_6\}$

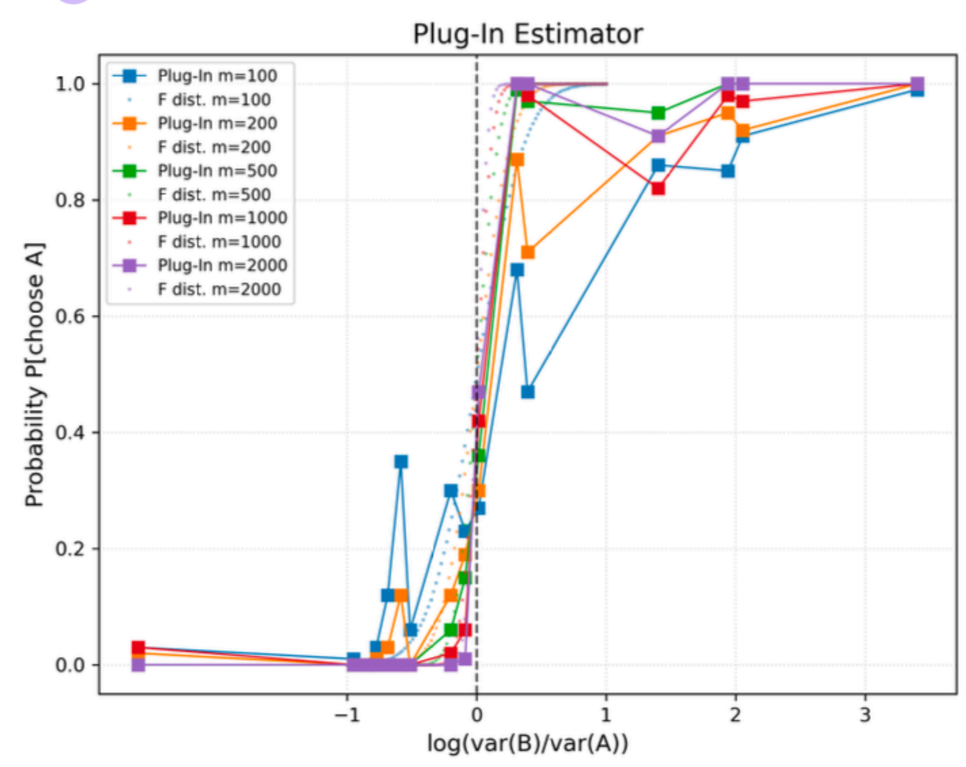


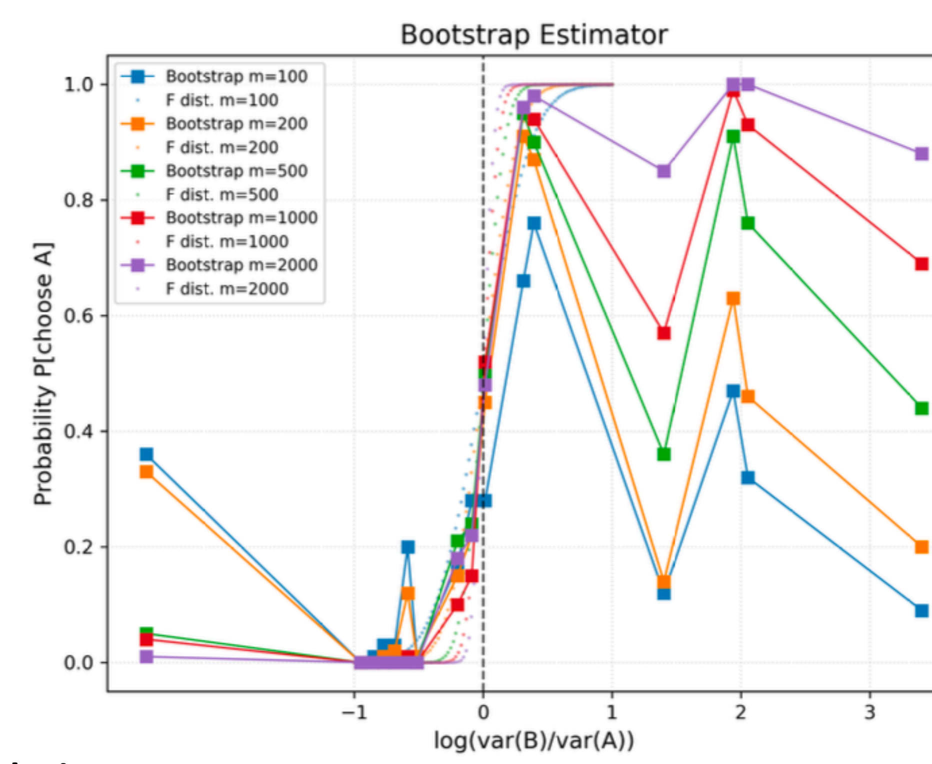




$A = \{V_4, V_5\}$

$B = \{V_6, V_7\}$





Conclusion:

- Plug-In estimator is more computationally efficient and accurate