A reconciliation between finite-sample and asymptopia-based methods in conditional independence testing

Ziang Niu

Department of Statistics, Wharton School

Aug. 10, 2023

- This a joint work with

- ArXiv link: https://arxiv.org/pdf/2211.14698.pdf

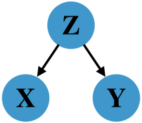
Outline

1. Hardness of CI Testing and regularity conditions $\mathscr{R}_n$
2. Two choices of $\mathscr{R}_n$: dCRT statistic and GCM statistic
3. $\widehat{\mathrm{dCRT}}$ Test and its equivalence to GCM Test
4. Numerical simulation

Outline

1. Hardness of CI Testing and regularity conditions $\mathscr{R}_n$
2. Two choices of $\mathscr{R}_n$: dCRT statistic and GCM statistic
3. $\widehat{\mathrm{dCRT}}$ Test and its equivalence to GCM Test
4. Numerical simulation

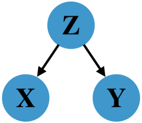
Hardness of conditional independence testing



- **Statistical task**: Consider the joint distribution $\mathscr{L}_n(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$, test the null hypothesis of conditional independence (CI):

$$H_0^{CI} = \boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Y} | \boldsymbol{Z}.$$

Hardness of conditional independence testing

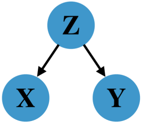


- **Statistical task**: Consider the joint distribution $\mathscr{L}_n(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$, test the null hypothesis of conditional independence (CI):

$$H_0^{CI} = \boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Y} | \boldsymbol{Z}.$$

- **Hardness of CI test**: According to Shah and Peters [2020],

Hardness of conditional independence testing



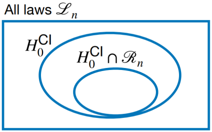
- **Statistical task**: Consider the joint distribution $\mathscr{L}_n(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$, test the null hypothesis of conditional independence (CI):

$$H_0^{CI} = \boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Y} | \boldsymbol{Z}.$$

- **Hardness of CI test**: According to Shah and Peters [2020],
  - If $\boldsymbol{Z}$ is continuous, any test with Type-I error control over the entire CI null $H_0^{CI} : \boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Y} | \boldsymbol{Z}$ cannot have nontrivial power against any alternative.
    $\Rightarrow$ a test with type-I error control must protect against too many sneaky ways $\boldsymbol{Z}$ can affect both $\boldsymbol{X}$ and $\boldsymbol{Y}$.

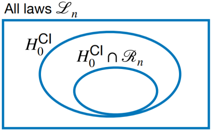
CI testing requires assumptions

- Given a set of regularity conditions $\mathscr{R}_n$ on $\mathscr{L}_n$, one can only hope to control Type-I error over the smaller null hypothesis: $H_0 : H_0^{CI} \cap \mathscr{R}_n$.



CI testing requires assumptions

- Given a set of regularity conditions $\mathscr{R}_n$ on $\mathscr{L}_n$, one can only hope to control Type-I error over the smaller null hypothesis: $H_0 : H_0^{CI} \cap \mathscr{R}_n$.



What kind of regularity conditions $\mathscr{R}_n$ should we impose?

Outline

1. Hardness of CI Testing and regularity conditions $\mathscr{R}_n$
2. Two choices of $\mathscr{R}_n$: dCRT statistic and GCM statistic
3. $\widehat{\mathrm{dCRT}}$ Test and its equivalence to GCM Test
4. Numerical simulation

dCRT statistic

dCRT statistic

- **Model-X (MX) assumption Candes et al. [2018]**: Assume we know the conditional distribution $\mathscr{L}_n(\boldsymbol{X}|\boldsymbol{Z})$ exactly, i.e.

$$\mathscr{R}_n = \left\{\mathscr{L}_n : \mathscr{L}_n(\boldsymbol{X}|\boldsymbol{Z}) = \mathscr{L}_n^*(\boldsymbol{X}|\boldsymbol{Z})\right\}.$$

where $\mathscr{L}_n^*(\boldsymbol{X}|\boldsymbol{Z})$ is the given conditional distribution.

dCRT statistic

- **Model-X (MX) assumption Candes et al. [2018]**: Assume we know the conditional distribution $\mathscr{L}_n(\boldsymbol{X}|\boldsymbol{Z})$ exactly, i.e.

$$\mathscr{R}_n = \left\{\mathscr{L}_n : \mathscr{L}_n(\boldsymbol{X}|\boldsymbol{Z}) = \mathscr{L}_n^*(\boldsymbol{X}|\boldsymbol{Z})\right\}.$$

  where $\mathscr{L}_n^*(\boldsymbol{X}|\boldsymbol{Z})$ is the given conditional distribution.

- MX assumption is reasonable if $\mathscr{L}_n(\boldsymbol{X}|\boldsymbol{Z})$ is controlled by experimenter. Powerful CI test are available under MX assumption (the conditional randomization test (CRT, Candes et al. [2018])).

dCRT statistic

- **Model-X (MX) assumption Candes et al. [2018]**: Assume we know the conditional distribution $\mathscr{L}_n(\boldsymbol{X}|\boldsymbol{Z})$ exactly, i.e.

$$\mathscr{R}_n = \{\mathscr{L}_n : \mathscr{L}_n(\boldsymbol{X}|\boldsymbol{Z}) = \mathscr{L}_n^*(\boldsymbol{X}|\boldsymbol{Z})\}.$$

  where $\mathscr{L}_n^*(\boldsymbol{X}|\boldsymbol{Z})$ is the given conditional distribution.
- MX assumption is reasonable if $\mathscr{L}_n(\boldsymbol{X}|\boldsymbol{Z})$ is controlled by experimenter. Powerful CI test are available under MX assumption (the conditional randomization test (CRT, Candes et al. [2018])).
- **A computationally more efficient way: dCRT Liu et al. [2022]**:

dCRT statistic

- **Model-X (MX) assumption Candes et al. [2018]**: Assume we know the conditional distribution $\mathscr{L}_n(\boldsymbol{X}|\boldsymbol{Z})$ exactly, i.e.

$$\mathscr{R}_n = \left\{\mathscr{L}_n : \mathscr{L}_n(\boldsymbol{X}|\boldsymbol{Z}) = \mathscr{L}_n^*(\boldsymbol{X}|\boldsymbol{Z})\right\}.$$

  where $\mathscr{L}_n^*(\boldsymbol{X}|\boldsymbol{Z})$ is the given conditional distribution.
- MX assumption is reasonable if $\mathscr{L}_n(\boldsymbol{X}|\boldsymbol{Z})$ is controlled by experimenter. Powerful CI test are available under MX assumption (the conditional randomization test (CRT, Candes et al. [2018])).
- **A computationally more efficient way: dCRT Liu et al. [2022]**:
  - Fit an approximation $\hat{\mu}_{n,y}(\boldsymbol{Z})$ of $\mu_{n,y}(\boldsymbol{Z}) = \mathbb{E}_{\mathscr{L}_n}[\boldsymbol{Y}|\boldsymbol{Z}]$ via machine learning;

dCRT statistic

- **Model-X (MX) assumption Candes et al. [2018]**: Assume we know the conditional distribution $\mathscr{L}_n(\boldsymbol{X}|\boldsymbol{Z})$ exactly, i.e.

$$\mathscr{R}_n = \left\{\mathscr{L}_n : \mathscr{L}_n(\boldsymbol{X}|\boldsymbol{Z}) = \mathscr{L}_n^*(\boldsymbol{X}|\boldsymbol{Z})\right\}.$$

  where $\mathscr{L}_n^*(\boldsymbol{X}|\boldsymbol{Z})$ is the given conditional distribution.

- MX assumption is reasonable if $\mathscr{L}_n(\boldsymbol{X}|\boldsymbol{Z})$ is controlled by experimenter. Powerful CI test are available under MX assumption (the conditional randomization test (CRT, Candes et al. [2018])).
- **A computationally more efficient way: dCRT Liu et al. [2022]**:
  - Fit an approximation $\hat{\mu}_{n,y}(\boldsymbol{Z})$ of $\mu_{n,y}(\boldsymbol{Z}) = \mathbb{E}_{\mathscr{L}_n}[\boldsymbol{Y}|\boldsymbol{Z}]$ via machine learning;
  - Compute $T_n(X, Y, Z) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu_{n,x}(Z_i))(Y_i - \hat{\mu}_{n,y}(Z_i))$;

dCRT statistic

- **Model-X (MX) assumption Candes et al. [2018]**: Assume we know the conditional distribution $\mathscr{L}_n(\boldsymbol{X}|\boldsymbol{Z})$ exactly, i.e.

$$\mathscr{R}_n = \left\{\mathscr{L}_n : \mathscr{L}_n(\boldsymbol{X}|\boldsymbol{Z}) = \mathscr{L}_n^*(\boldsymbol{X}|\boldsymbol{Z})\right\}.$$

  where $\mathscr{L}_n^*(\boldsymbol{X}|\boldsymbol{Z})$ is the given conditional distribution.
- MX assumption is reasonable if $\mathscr{L}_n(\boldsymbol{X}|\boldsymbol{Z})$ is controlled by experimenter. Powerful CI test are available under MX assumption (the conditional randomization test (CRT, Candes et al. [2018])).
- **A computationally more efficient way: dCRT Liu et al. [2022]**:
  - Fit an approximation $\hat{\mu}_{n,y}(\boldsymbol{Z})$ of $\mu_{n,y}(\boldsymbol{Z}) = \mathbb{E}_{\mathscr{L}_n}[\boldsymbol{Y}|\boldsymbol{Z}]$ via machine learning;
  - Compute $T_n(X,Y,Z) = \frac{1}{\sqrt{n}}\sum_{i=1}^n (X_i - \mu_{n,x}(Z_i))(Y_i - \hat{\mu}_{n,y}(Z_i))$;
  - For $b = 1, \ldots, B$

dCRT statistic

- **Model-X (MX) assumption Candes et al. [2018]**: Assume we know the conditional distribution $\mathscr{L}_n(\boldsymbol{X}|\boldsymbol{Z})$ exactly, i.e.

$$\mathscr{R}_n = \left\{\mathscr{L}_n : \mathscr{L}_n(\boldsymbol{X}|\boldsymbol{Z}) = \mathscr{L}_n^*(\boldsymbol{X}|\boldsymbol{Z})\right\}.$$

  where $\mathscr{L}_n^*(\boldsymbol{X}|\boldsymbol{Z})$ is the given conditional distribution.
- MX assumption is reasonable if $\mathscr{L}_n(\boldsymbol{X}|\boldsymbol{Z})$ is controlled by experimenter. Powerful CI test are available under MX assumption (the conditional randomization test (CRT, Candes et al. [2018])).
- **A computationally more efficient way: dCRT Liu et al. [2022]**:
  - Fit an approximation $\hat{\mu}_{n,y}(\boldsymbol{Z})$ of $\mu_{n,y}(\boldsymbol{Z}) = \mathbb{E}_{\mathscr{L}_n}[\boldsymbol{Y}|\boldsymbol{Z}]$ via machine learning;
  - Compute $T_n(X,Y,Z) = \frac{1}{\sqrt{n}}\sum_{i=1}^n (X_i - \mu_{n,x}(Z_i))(Y_i - \hat{\mu}_{n,y}(Z_i))$;
  - For $b = 1, \ldots, B$
    - Draw $\tilde{X}_i^{(b)} \sim \mathscr{L}_n^*(X_i|\boldsymbol{Z} = Z_i)$;

dCRT statistic

- **Model-X (MX) assumption Candes et al. [2018]**: Assume we know the conditional distribution $\mathscr{L}_n(\boldsymbol{X}|\boldsymbol{Z})$ exactly, i.e.

$$\mathscr{R}_n = \left\{\mathscr{L}_n : \mathscr{L}_n(\boldsymbol{X}|\boldsymbol{Z}) = \mathscr{L}_n^*(\boldsymbol{X}|\boldsymbol{Z})\right\}.$$

  where $\mathscr{L}_n^*(\boldsymbol{X}|\boldsymbol{Z})$ is the given conditional distribution.
- MX assumption is reasonable if $\mathscr{L}_n(\boldsymbol{X}|\boldsymbol{Z})$ is controlled by experimenter. Powerful CI test are available under MX assumption (the conditional randomization test (CRT, Candes et al. [2018])).
- **A computationally more efficient way: dCRT Liu et al. [2022]**:
  - Fit an approximation $\hat{\mu}_{n,y}(\boldsymbol{Z})$ of $\mu_{n,y}(\boldsymbol{Z}) = \mathbb{E}_{\mathscr{L}_n}[\boldsymbol{Y}|\boldsymbol{Z}]$ via machine learning;
  - Compute $T_n(X,Y,Z) = \frac{1}{\sqrt{n}}\sum_{i=1}^n (X_i - \mu_{n,x}(Z_i))(Y_i - \hat{\mu}_{n,y}(Z_i))$;
  - For $b = 1, \ldots, B$
    - Draw $\tilde{X}_i^{(b)} \sim \mathscr{L}_n^*(X_i|\boldsymbol{Z} = Z_i)$;
    - Compute test statistic

$$T_n(\tilde{X}^{(b)}, X, Y, Z) = \frac{1}{\sqrt{n}}\sum_{i=1}^n (\tilde{X}_i^{(b)} - \mu_{n,x}(Z_i))(Y_i - \hat{\mu}_{n,y}(Z_i));$$

dCRT statistic

- **Model-X (MX) assumption Candes et al. [2018]**: Assume we know the conditional distribution $\mathscr{L}_n(\boldsymbol{X}|\boldsymbol{Z})$ exactly, i.e.

$$\mathscr{R}_n = \left\{\mathscr{L}_n : \mathscr{L}_n(\boldsymbol{X}|\boldsymbol{Z}) = \mathscr{L}_n^*(\boldsymbol{X}|\boldsymbol{Z})\right\}.$$

  where $\mathscr{L}_n^*(\boldsymbol{X}|\boldsymbol{Z})$ is the given conditional distribution.
- MX assumption is reasonable if $\mathscr{L}_n(\boldsymbol{X}|\boldsymbol{Z})$ is controlled by experimenter. Powerful CI test are available under MX assumption (the conditional randomization test (CRT, Candes et al. [2018])).
- **A computationally more efficient way: dCRT Liu et al. [2022]**:
  - Fit an approximation $\hat{\mu}_{n,y}(\boldsymbol{Z})$ of $\mu_{n,y}(\boldsymbol{Z}) = \mathbb{E}_{\mathscr{L}_n}[\boldsymbol{Y}|\boldsymbol{Z}]$ via machine learning;
  - Compute $T_n(X, Y, Z) = \frac{1}{\sqrt{n}}\sum_{i=1}^n (X_i - \mu_{n,x}(Z_i))(Y_i - \hat{\mu}_{n,y}(Z_i))$;
  - For $b = 1, \ldots, B$
    - Draw $\tilde{X}_i^{(b)} \sim \mathscr{L}_n^*(X_i|\boldsymbol{Z} = Z_i)$;
    - Compute test statistic

$$T_n(\tilde{X}^{(b)}, X, Y, Z) = \frac{1}{\sqrt{n}}\sum_{i=1}^n (\tilde{X}_i^{(b)} - \mu_{n,x}(Z_i))(Y_i - \hat{\mu}_{n,y}(Z_i));$$

    - Compute $C_{n,\alpha}(X, Y, Z) = \mathbb{Q}_{1-\alpha}[\{T_n(X, Y, Z), T_n(\tilde{X}^{(1)}, X, Y, Z), \ldots, T_n(\tilde{X}^B, X, Y, Z)\}]$.

dCRT statistic

- **Model-X (MX) assumption Candes et al. [2018]**: Assume we know the conditional distribution $\mathscr{L}_n(\boldsymbol{X}|\boldsymbol{Z})$ exactly, i.e.

$$\mathscr{R}_n = \left\{\mathscr{L}_n : \mathscr{L}_n(\boldsymbol{X}|\boldsymbol{Z}) = \mathscr{L}_n^*(\boldsymbol{X}|\boldsymbol{Z})\right\}.$$

  where $\mathscr{L}_n^*(\boldsymbol{X}|\boldsymbol{Z})$ is the given conditional distribution.
- MX assumption is reasonable if $\mathscr{L}_n(\boldsymbol{X}|\boldsymbol{Z})$ is controlled by experimenter. Powerful CI test are available under MX assumption (the conditional randomization test (CRT, Candes et al. [2018])).
- **A computationally more efficient way: dCRT Liu et al. [2022]**:
  - Fit an approximation $\hat{\mu}_{n,y}(\boldsymbol{Z})$ of $\mu_{n,y}(\boldsymbol{Z}) = \mathbb{E}_{\mathscr{L}_n}[\boldsymbol{Y}|\boldsymbol{Z}]$ via machine learning;
  - Compute $T_n(X, Y, Z) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mu_{n,x}(Z_i))(Y_i - \hat{\mu}_{n,y}(Z_i))$;
  - For $b = 1, \ldots, B$
    - Draw $\tilde{X}_i^{(b)} \sim \mathscr{L}_n^*(X_i|\boldsymbol{Z} = Z_i)$;
    - Compute test statistic

$$T_n(\tilde{X}^{(b)}, X, Y, Z) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\tilde{X}_i^{(b)} - \mu_{n,x}(Z_i))(Y_i - \hat{\mu}_{n,y}(Z_i));$$

    - Compute $C_{n,\alpha}(X, Y, Z) = \mathbb{Q}_{1-\alpha}[\{T_n(X, Y, Z), T_n(\tilde{X}^{(1)}, X, Y, Z), \ldots, T_n(\tilde{X}^B, X, Y, Z)\}]$.
    - Reject if $T_n(X, Y, Z) > C_{n,\alpha}(X, Y, Z)$.

Doubly robust conditional independence test: GCM test

- **The GCM test Shah and Peters [2020]**:

Doubly robust conditional independence test: GCM test

- **The GCM test Shah and Peters [2020]**:
  - Fit an approximation $\hat{\mu}_{n,x}(\boldsymbol{Z})$ of $\mu_{n,x}(\boldsymbol{Z}) = \mathbb{E}_{\mathscr{L}_n}[\boldsymbol{X}|\boldsymbol{Z}]$ via machine learning;

Doubly robust conditional independence test: GCM test

- **The GCM test Shah and Peters [2020]**:
  - Fit an approximation $\hat{\mu}_{n,x}(\boldsymbol{Z})$ of $\mu_{n,x}(\boldsymbol{Z}) = \mathbb{E}_{\mathscr{L}_n}[\boldsymbol{X}|\boldsymbol{Z}]$ via machine learning;
  - Fit an approximation $\hat{\mu}_{n,y}(\boldsymbol{Z})$ of $\mu_{n,y}(\boldsymbol{Z}) = \mathbb{E}_{\mathscr{L}_n}[\boldsymbol{Y}|\boldsymbol{Z}]$ via machine learning;

Doubly robust conditional independence test: GCM test

- **The GCM test Shah and Peters [2020]**:
  - Fit an approximation $\hat{\mu}_{n,x}(\boldsymbol{Z})$ of $\mu_{n,x}(\boldsymbol{Z}) = \mathbb{E}_{\mathscr{L}_n}[\boldsymbol{X}|\boldsymbol{Z}]$ via machine learning;
  - Fit an approximation $\hat{\mu}_{n,y}(\boldsymbol{Z})$ of $\mu_{n,y}(\boldsymbol{Z}) = \mathbb{E}_{\mathscr{L}_n}[\boldsymbol{Y}|\boldsymbol{Z}]$ via machine learning;
  - Compute $T_n(X, Y, Z) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \hat{\mu}_{n,x}(Z_i))(Y_i - \hat{\mu}_{n,y}(Z_i))$;

Doubly robust conditional independence test: GCM test

- **The GCM test Shah and Peters [2020]**:
  - Fit an approximation $\hat{\mu}_{n,x}(\boldsymbol{Z})$ of $\mu_{n,x}(\boldsymbol{Z}) = \mathbb{E}_{\mathscr{L}_n}[\boldsymbol{X}|\boldsymbol{Z}]$ via machine learning;
  - Fit an approximation $\hat{\mu}_{n,y}(\boldsymbol{Z})$ of $\mu_{n,y}(\boldsymbol{Z}) = \mathbb{E}_{\mathscr{L}_n}[\boldsymbol{Y}|\boldsymbol{Z}]$ via machine learning;
  - Compute $T_n(X, Y, Z) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \hat{\mu}_{n,x}(Z_i))(Y_i - \hat{\mu}_{n,y}(Z_i))$;
  - Compute $(S_n^{GCM})^2(X, Y, Z) = \mathrm{Var}[(X_i - \hat{\mu}_{n,x}(Z_i))(Y_i - \hat{\mu}_{n,y}(Z_i))]$ (sample variance);

Doubly robust conditional independence test: GCM test

- **The GCM test Shah and Peters [2020]**:
  - Fit an approximation $\hat{\mu}_{n,x}(\boldsymbol{Z})$ of $\mu_{n,x}(\boldsymbol{Z}) = \mathbb{E}_{\mathscr{L}_n}[\boldsymbol{X}|\boldsymbol{Z}]$ via machine learning;
  - Fit an approximation $\hat{\mu}_{n,y}(\boldsymbol{Z})$ of $\mu_{n,y}(\boldsymbol{Z}) = \mathbb{E}_{\mathscr{L}_n}[\boldsymbol{Y}|\boldsymbol{Z}]$ via machine learning;
  - Compute $T_n(X, Y, Z) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \hat{\mu}_{n,x}(Z_i))(Y_i - \hat{\mu}_{n,y}(Z_i))$;
  - Compute $(S_n^{GCM})^2(X, Y, Z) = \mathrm{Var}[(X_i - \hat{\mu}_{n,x}(Z_i))(Y_i - \hat{\mu}_{n,y}(Z_i))]$ (sample variance);
  - Reject if $T_n(X, Y, Z)/S_n^{GCM}(X, Y, Z) > z_{1-\alpha}$.

Doubly robust conditional independence test: GCM test

- **The GCM test Shah and Peters [2020]**:
  - Fit an approximation $\hat{\mu}_{n,x}(\boldsymbol{Z})$ of $\mu_{n,x}(\boldsymbol{Z}) = \mathbb{E}_{\mathscr{L}_n}[\boldsymbol{X}|\boldsymbol{Z}]$ via machine learning;
  - Fit an approximation $\hat{\mu}_{n,y}(\boldsymbol{Z})$ of $\mu_{n,y}(\boldsymbol{Z}) = \mathbb{E}_{\mathscr{L}_n}[\boldsymbol{Y}|\boldsymbol{Z}]$ via machine learning;
  - Compute $T_n(X,Y,Z) = \frac{1}{\sqrt{n}}\sum_{i=1}^n (X_i - \hat{\mu}_{n,x}(Z_i))(Y_i - \hat{\mu}_{n,y}(Z_i))$;
  - Compute $(S_n^{GCM})^2(X,Y,Z) = \mathrm{Var}[(X_i - \hat{\mu}_{n,x}(Z_i))(Y_i - \hat{\mu}_{n,y}(Z_i))]$ (sample variance);
  - Reject if $T_n(X,Y,Z)/S_n^{GCM}(X,Y,Z) > z_{1-\alpha}$.
- This is an asymptopia-based method as opposed to resampling nature of dCRT. It enjoys the double robustness.

Theorem 1 (Shah and Peters [2020]; informal)

For $\mathscr{L}_n \in H_0 = H_0^{CI} \cap \mathscr{R}_n$, where $\mathscr{R}_n$ is defined as a set of laws satisfying

$$\left\{\mathrm{RMSE}(\hat{\mu}_{n,x}) = o_P(1), \mathrm{RMSE}(\hat{\mu}_{n,y}) = o_P(1), \mathrm{RMSE}(\hat{\mu}_{n,x}) \cdot \mathrm{RMSE}(\hat{\mu}_{n,y}) = o_P(n^{-1/2})\right\},$$

then we have

$$\limsup_{n\to\infty} \sup_{\mathscr{L}_n \in H_0} \mathbb{P}_{\mathscr{L}_n}[\textit{GCM rejects null}\,] \leq \alpha.$$

Outline

Recall dCRT statistic

- **MX framework: dCRT Liu et al. [2022]**:
  - Fit an approximation $\hat{\mu}_{n,y}(\boldsymbol{Z})$ of $\mu_{n,y}(\boldsymbol{Z}) = \mathbb{E}_{\mathscr{L}_n}[\boldsymbol{Y}|\boldsymbol{Z}]$ via machine learning;
  - Compute $T_n(X, Y, Z) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu_{n,x}(Z_i))(Y_i - \hat{\mu}_{n,y}(Z_i))$;
  - For $b = 1, \ldots, B$
    - Draw $\tilde{X}_i^{(b)} \sim \mathscr{L}_n^*(X_i|\boldsymbol{Z} = Z_i)$;
    - Compute test statistic

$$T_n(\tilde{X}^{(b)}, X, Y, Z) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\tilde{X}_i^{(b)} - \mu_{n,x}(Z_i))(Y_i - \hat{\mu}_{n,y}(Z_i));$$

    - Compute

$$C_{n,\alpha}(X, Y, Z) = \mathbb{Q}_{1-\alpha}[\{T_n(X, Y, Z), T_n(\tilde{X}^{(1)}, X, Y, Z), \ldots, T_n(\tilde{X}^{B}, X, Y, Z)\}].$$

    - Reject if $T_n(X, Y, Z) > C_{n,\alpha}(X, Y, Z)$.

Recall dCRT statistic

- **MX framework: dCRT Liu et al. [2022]**:
  - Fit an approximation $\hat{\mu}_{n,y}(\boldsymbol{Z})$ of $\mu_{n,y}(\boldsymbol{Z}) = \mathbb{E}_{\mathscr{L}_n}[\boldsymbol{Y}|\boldsymbol{Z}]$ via machine learning;
  - Compute $T_n(X, Y, Z) = \frac{1}{\sqrt{n}}\sum_{i=1}^n (X_i - \mu_{n,x}(Z_i))(Y_i - \hat{\mu}_{n,y}(Z_i))$;
  - For $b = 1, \ldots, B$
    - Draw $\tilde{X}_i^{(b)} \sim \mathscr{L}_n^*(X_i|\boldsymbol{Z} = Z_i)$;
    - Compute test statistic

$$T_n(\tilde{X}^{(b)}, X, Y, Z) = \frac{1}{\sqrt{n}}\sum_{i=1}^n (\tilde{X}_i^{(b)} - \mu_{n,x}(Z_i))(Y_i - \hat{\mu}_{n,y}(Z_i));$$

    - Compute

$$C_{n,\alpha}(X, Y, Z) = \mathbb{Q}_{1-\alpha}[\{T_n(X, Y, Z), T_n(\tilde{X}^{(1)}, X, Y, Z), \ldots, T_n(\tilde{X}^{B}, X, Y, Z)\}].$$

    - Reject if $T_n(X, Y, Z) > C_{n,\alpha}(X, Y, Z)$.

- **Challenge**: $\mathscr{L}_n^*(\boldsymbol{X}|\boldsymbol{Z})$ is usually an approximation in practice!

Recall dCRT statistic

- **MX framework: dCRT Liu et al. [2022]**:
  - Fit an approximation $\hat{\mu}_{n,y}(\boldsymbol{Z})$ of $\mu_{n,y}(\boldsymbol{Z}) = \mathbb{E}_{\mathcal{L}_n}[\boldsymbol{Y}|\boldsymbol{Z}]$ via machine learning;
  - Compute $T_n(X, Y, Z) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu_{n,x}(Z_i))(Y_i - \hat{\mu}_{n,y}(Z_i))$;
  - For $b = 1, \ldots, B$
    - Draw $\tilde{X}_i^{(b)} \sim \mathcal{L}_n^*(X_i|\boldsymbol{Z} = Z_i)$;
    - Compute test statistic

$$T_n(\tilde{X}^{(b)}, X, Y, Z) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\tilde{X}_i^{(b)} - \mu_{n,x}(Z_i))(Y_i - \hat{\mu}_{n,y}(Z_i));$$

    - Compute

$$C_{n,\alpha}(X, Y, Z) = \mathbb{Q}_{1-\alpha}[\{T_n(X, Y, Z), T_n(\tilde{X}^{(1)}, X, Y, Z), \ldots, T_n(\tilde{X}^{B}, X, Y, Z)\}].$$

    - Reject if $T_n(X, Y, Z) > C_{n,\alpha}(X, Y, Z)$.
- **Challenge**: $\mathcal{L}_n^*(\boldsymbol{X}|\boldsymbol{Z})$ is usually an approximation in practice!
- **Our focus**: Robustness and power of MX (dCRT) methods when $\mathcal{L}_n^*(\boldsymbol{X}|\boldsymbol{Z})$ learned in sample. In other words, replace $\mu_{n,x}(\cdot)$ with the estimate $\hat{\mu}_{n,x}(\cdot)$ and draw resamples from the learned distribution $\hat{\mathcal{L}}_n^*(X_i|\boldsymbol{Z} = Z_i)$.

$\overline{\text{dCRT}}$ statistic

Procedure:

- Fit an approximation $\hat{\mu}_{n,y}(\boldsymbol{Z})$ of $\mu_{n,y}(\boldsymbol{Z}) = \mathbb{E}_{\mathcal{L}_n}[\boldsymbol{Y}|\boldsymbol{Z}]$ via machine learning;
- Fit an approximation $\hat{\mu}_{n,x}(\boldsymbol{Z})$ of $\mu_{n,x}(\boldsymbol{Z}) = \mathbb{E}_{\mathcal{L}_n}[\boldsymbol{X}|\boldsymbol{Z}]$ via machine learning;
- Compute $T_n(X, Y, Z) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mu_{n,x}(Z_i))(Y_i - \hat{\mu}_{n,y}(Z_i))$;
- For $b = 1, \ldots, B$
  - Draw $\tilde{X}_i^{(b)} \sim \hat{\mathcal{L}}_n^*(X_i|\boldsymbol{Z} = Z_i)$;
  - Compute test statistic

$$T_n(\tilde{X}^{(b)}, X, Y, Z) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\tilde{X}_i^{(b)} - \hat{\mu}_{n,x}(Z_i))(Y_i - \hat{\mu}_{n,y}(Z_i));$$

  - Compute

$$C_{n,\alpha}(X, Y, Z) = \mathbb{Q}_{1-\alpha}[\{T_n(X, Y, Z), T_n(\tilde{X}^{(1)}, X, Y, Z), \ldots, T_n(\tilde{X}^{B}, X, Y, Z)\}].$$

  - Reject if $T_n(X, Y, Z) > C_{n,\alpha}(X, Y, Z)$.

Comparison between $\widehat{\mathrm{dCRT}}$ to GCM test

- Recall the test statistic and resampling test statistic

$$T_n(X, Y, Z) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \hat{\mu}_{n,x})(Y_i - \hat{\mu}_{n,y}(Z_i))$$

$$T_n(\tilde{X}, X, Y, Z) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\tilde{X}_i - \hat{\mu}_{n,x}(Z_i))(Y_i - \hat{\mu}_{n,y}(Z_i)).$$

Comparison between $\widehat{\mathrm{dCRT}}$ to GCM test

- Recall the test statistic and resampling test statistic

$$T_n(X,Y,Z)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-\hat{\mu}_{n,x})(Y_i-\hat{\mu}_{n,y}(Z_i))$$

$$T_n(\tilde{X},X,Y,Z)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\tilde{X}_i-\hat{\mu}_{n,x}(Z_i))(Y_i-\hat{\mu}_{n,y}(Z_i)).$$

- $\widehat{\mathrm{dCRT}}$ rejects if $T_n(X,Y,Z)>\mathbb{Q}_{1-\alpha}[T_n(\tilde{X},X,Y,Z)|X,Y,Z]$ or equivalently if

$$\frac{T_n(X,Y,Z)}{S_n^{\widehat{\mathrm{dCRT}}}(X,Y,Z)}>\mathbb{Q}_{1-\alpha}\left[\frac{T_n(\tilde{X},X,Y,Z)}{S_n^{\widehat{\mathrm{dCRT}}}(X,Y,Z)}|X,Y,Z\right],$$

$$S_n^{\widehat{\mathrm{dCRT}}}(X,Y,Z)=\frac{1}{n}\sum_{i=1}^{n}\mathrm{Var}_{\hat{\mathcal{L}}_n}[X_i|Z_i](Y_i-\hat{\mu}_{n,y}(Z_i))^2.$$

Comparison between $\widehat{\mathrm{dCRT}}$ to GCM test

- Recall the test statistic and resampling test statistic

$$T_n(X, Y, Z) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \hat{\mu}_{n,x})(Y_i - \hat{\mu}_{n,y}(Z_i))$$

$$T_n(\tilde{X}, X, Y, Z) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\tilde{X}_i - \hat{\mu}_{n,x}(Z_i))(Y_i - \hat{\mu}_{n,y}(Z_i)).$$

- $\widehat{\mathrm{dCRT}}$ rejects if $T_n(X, Y, Z) > \mathbb{Q}_{1-\alpha}[T_n(\tilde{X}, X, Y, Z)|X, Y, Z]$ or equivalently if

$$\frac{T_n(X, Y, Z)}{S_n^{\widehat{\mathrm{dCRT}}}(X, Y, Z)} > \mathbb{Q}_{1-\alpha}\left[\frac{T_n(\tilde{X}, X, Y, Z)}{S_n^{\widehat{\mathrm{dCRT}}}(X, Y, Z)}|X, Y, Z\right],$$

$$S_n^{\widehat{\mathrm{dCRT}}}(X, Y, Z) = \frac{1}{n} \sum_{i=1}^{n} \mathrm{Var}_{\hat{\mathscr{L}}_n}[X_i|Z_i](Y_i - \hat{\mu}_{n,y}(Z_i))^2.$$

- GCM rejects if $T_n(X, Y, Z)/S_n^{GCM}(X, Y, Z) > z_{1-\alpha}$.

Comparison between $\widehat{\mathrm{dCRT}}$ to GCM test

- Recall the test statistic and resampling test statistic

$$T_n(X,Y,Z) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - \hat{\mu}_{n,x})(Y_i - \hat{\mu}_{n,y}(Z_i))$$

$$T_n(\tilde{X},X,Y,Z) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\tilde{X}_i - \hat{\mu}_{n,x}(Z_i))(Y_i - \hat{\mu}_{n,y}(Z_i)).$$

- $\widehat{\mathrm{dCRT}}$ rejects if $T_n(X,Y,Z) > \mathbb{Q}_{1-\alpha}[T_n(\tilde{X},X,Y,Z)|X,Y,Z]$ or equivalently if

$$\frac{T_n(X,Y,Z)}{S_n^{\widehat{\mathrm{dCRT}}}(X,Y,Z)} > \mathbb{Q}_{1-\alpha}\left[\frac{T_n(\tilde{X},X,Y,Z)}{S_n^{\widehat{\mathrm{dCRT}}}(X,Y,Z)}|X,Y,Z\right],$$

$$S_n^{\widehat{\mathrm{dCRT}}}(X,Y,Z) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{Var}_{\hat{\mathscr{L}}_n}[X_i|Z_i](Y_i - \hat{\mu}_{n,y}(Z_i))^2.$$

- GCM rejects if $T_n(X,Y,Z)/S_n^{GCM}(X,Y,Z) > z_{1-\alpha}$.
- Two tests are asymptotic equivalent if

Comparison between $\widehat{\mathrm{dCRT}}$ to GCM test

- Recall the test statistic and resampling test statistic

$$T_n(X,Y,Z) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - \hat{\mu}_{n,x})(Y_i - \hat{\mu}_{n,y}(Z_i))$$

$$T_n(\tilde{X},X,Y,Z) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\tilde{X}_i - \hat{\mu}_{n,x}(Z_i))(Y_i - \hat{\mu}_{n,y}(Z_i)).$$

- $\widehat{\mathrm{dCRT}}$ rejects if $T_n(X,Y,Z) > \mathbb{Q}_{1-\alpha}[T_n(\tilde{X},X,Y,Z)|X,Y,Z]$ or equivalently if

$$\frac{T_n(X,Y,Z)}{S_n^{\widehat{\mathrm{dCRT}}}(X,Y,Z)} > \mathbb{Q}_{1-\alpha}\left[\frac{T_n(\tilde{X},X,Y,Z)}{S_n^{\widehat{\mathrm{dCRT}}}(X,Y,Z)}|X,Y,Z\right],$$

$$S_n^{\widehat{\mathrm{dCRT}}}(X,Y,Z) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{Var}_{\hat{\mathscr{L}}_n}[X_i|Z_i](Y_i - \hat{\mu}_{n,y}(Z_i))^2.$$

- GCM rejects if $T_n(X,Y,Z)/S_n^{GCM}(X,Y,Z) > z_{1-\alpha}$.
- Two tests are asymptotic equivalent if
  1. Variance estimates $S_n^{GCM}, S_n^{\widehat{\mathrm{dCRT}}}$ are asymptotically equivalent;

Comparison between $\widehat{\mathrm{dCRT}}$ to GCM test

- Recall the test statistic and resampling test statistic

$$T_n(X,Y,Z) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - \hat{\mu}_{n,x})(Y_i - \hat{\mu}_{n,y}(Z_i))$$

$$T_n(\tilde{X},X,Y,Z) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\tilde{X}_i - \hat{\mu}_{n,x}(Z_i))(Y_i - \hat{\mu}_{n,y}(Z_i)).$$

- $\widehat{\mathrm{dCRT}}$ rejects if $T_n(X,Y,Z) > \mathbb{Q}_{1-\alpha}[T_n(\tilde{X},X,Y,Z)|X,Y,Z]$ or equivalently if

$$\frac{T_n(X,Y,Z)}{S_n^{\widehat{\mathrm{dCRT}}}(X,Y,Z)} > \mathbb{Q}_{1-\alpha}\left[\frac{T_n(\tilde{X},X,Y,Z)}{S_n^{\widehat{\mathrm{dCRT}}}(X,Y,Z)}|X,Y,Z\right],$$

$$S_n^{\widehat{\mathrm{dCRT}}}(X,Y,Z) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{Var}_{\hat{\mathscr{L}}_n}[X_i|Z_i](Y_i - \hat{\mu}_{n,y}(Z_i))^2.$$

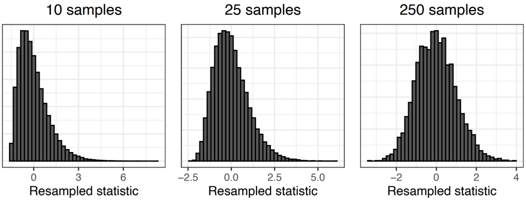
- GCM rejects if $T_n(X,Y,Z)/S_n^{GCM}(X,Y,Z) > z_{1-\alpha}$.
- Two tests are asymptotic equivalent if
  1. Variance estimates $S_n^{GCM}, S_n^{\widehat{\mathrm{dCRT}}}$ are asymptotically equivalent;
  2. The normalized $\widehat{\mathrm{dCRT}}$ resmapling distribution convergence to $N(0,1)$.

Convergence of $\widehat{\mathrm{dCRT}}$ resampling distribution to $N(0,1)$

- For small n, $\widehat{\mathrm{dCRT}}$ resampling distribution need not be normal; increasing n brings resampling distribution closer to normality.

Convergence of $\widehat{\text{dCRT}}$ resampling distribution to $N(0, 1)$

- For small n, $\widehat{\text{dCRT}}$ resampling distribution need not be normal; increasing n brings resampling distribution closer to normality.



$\widehat{\text{dCRT}}$-GCM equivalence and $\widehat{\text{dCRT}}$ robustness

Equivalence result:

Theorem (Niu et al '22; informal). Assume

1. RMSE($\hat{\mu}_{n,x}$) $= o_P(1)$, RMSE($\hat{\mu}_{n,y}$) $= o_P(1)$, RMSE($\hat{\mu}_{n,x}$) $\cdot$ RMSE($\hat{\mu}_{n,y}$) $= o_P(n^{-1/2})$.
2. The estimated variances are consistent in the following sense:

$$\frac{1}{n}\sum_{i=1}^{n}(\text{Var}_{\widehat{\mathscr{L}}_n}[X_i \mid Z_i] - \text{Var}_{\mathscr{L}_n}[X_i \mid Z_i])\text{Var}_{\mathscr{L}_n}[Y_i \mid Z_i] \xrightarrow{p} 0.$$

Then, for any $\mathscr{L}_n \in H_0$, the dCRT is asymptotically equivalent to the GCM test, i.e.

$$\lim_{n\to\infty} \inf_{\mathscr{L}_n \in H_0} \mathbb{P}_{\mathscr{L}_n}[\text{GCM test and dCRT coincide}] = 1.$$

$\widehat{\text{dCRT}}$-GCM equivalence and $\widehat{\text{dCRT}}$ robustness

Equivalence result:

Theorem (Niu et al '22; informal). Assume

1. $\text{RMSE}(\hat{\mu}_{n,x}) = o_P(1)$, $\text{RMSE}(\hat{\mu}_{n,y}) = o_P(1)$, $\text{RMSE}(\hat{\mu}_{n,x}) \cdot \text{RMSE}(\hat{\mu}_{n,y}) = o_P(n^{-1/2})$.
2. The estimated variances are consistent in the following sense:

$$\frac{1}{n}\sum_{i=1}^{n}(\text{Var}_{\widehat{\mathcal{L}}_n}[X_i \mid Z_i] - \text{Var}_{\mathcal{L}_n}[X_i \mid Z_i])\text{Var}_{\mathcal{L}_n}[Y_i \mid Z_i] \xrightarrow{p} 0.$$

Then, for any $\mathcal{L}_n \in H_0$, the dCRT is asymptotically equivalent to the GCM test, i.e.

$$\lim_{n\to\infty} \inf_{\mathcal{L}_n \in H_0} \mathbb{P}_{\mathcal{L}_n}[\text{GCM test and dCRT coincide}] = 1.$$

Double robustness result:

Corollary (Niu et al '22; informal). Given conditions 1 & 2, dCRT is doubly robust:

$$\limsup_{n\to\infty} \sup_{\mathcal{L}_n \in H_0} \mathbb{P}_{\mathcal{L}_n}[\text{dCRT rejects null}] \leq \alpha.$$

Outline

1. Hardness of CI Testing and regularity conditions $\mathscr{R}_n$
2. Two choices of $\mathscr{R}_n$: dCRT statistic and GCM statistic
3. $\widehat{\mathrm{dCRT}}$ Test and its equivalence to GCM Test
4. Numerical simulation

Numerical simulation: Design

- Consider

$$Z \sim N(0, \Sigma(p)),\ \mathscr{L}(X|Z) = N(Z^\top \beta, 1),\ \mathscr{L}(Y|X,Z) = N(X\theta + Z^\top \beta, 1)$$

where

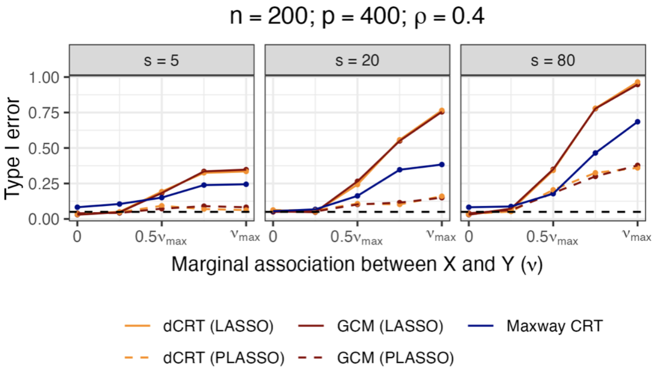
$$\Sigma_{ij}(p) = \rho^{|i-j|},\ \beta_j = \begin{cases} \nu & \text{if } j \leq s, \\ 0 & \text{if } j > s. \end{cases}$$

Parameters ν and θ control degree of confounding and signal strength.

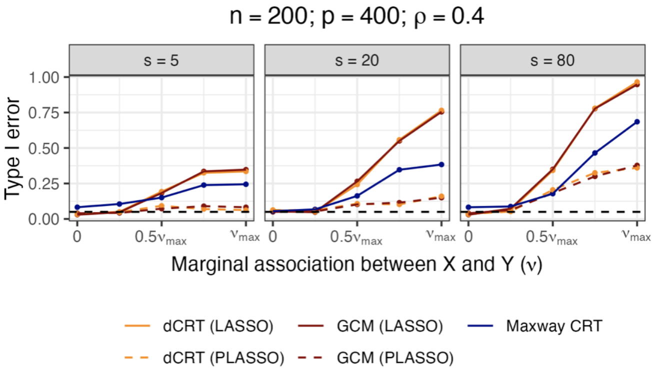
- Methods compared:
  - $\widehat{\mathrm{dCRT}}$[1] and GCM (with lasso and post-lasso);
  - Maxway CRT (a competitive method).

[1]We will use dCRT instead in the following.

Numerical simulations: Type-I error control

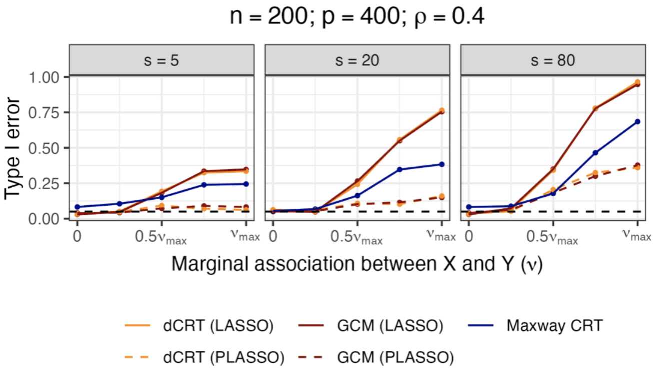


Numerical simulations: Type-I error control



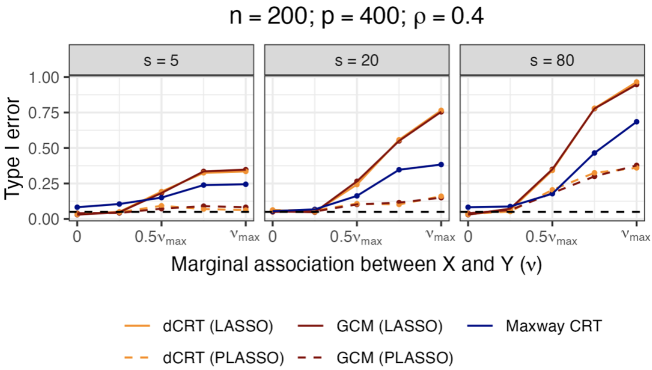
- Some takeaways:

Numerical simulations: Type-I error control



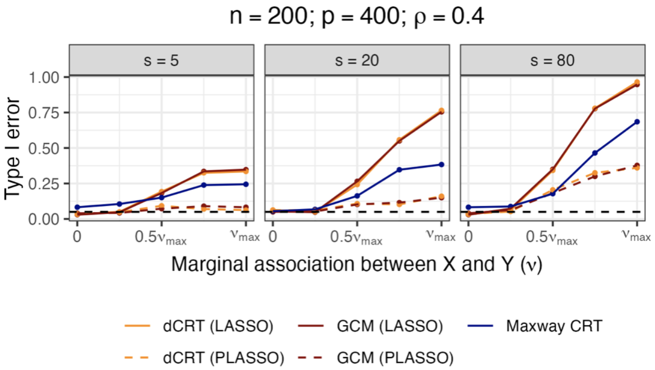
- Some takeaways:
  - GCM and dCRT perform similarly, consistent with asymptotic theory.

Numerical simulations: Type-I error control



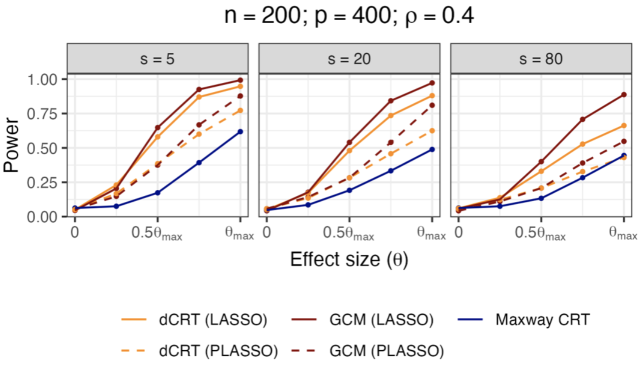
- Some takeaways:
  - GCM and dCRT perform similarly, consistent with asymptotic theory.
  - Lasso-based methods can have very inflated Type-I error in difficult settings.

Numerical simulations: Type-I error control

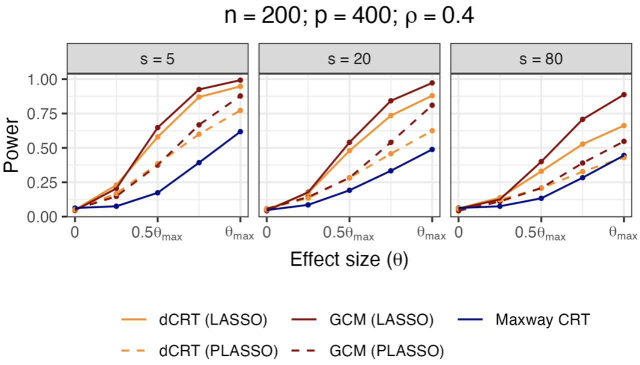


- Some takeaways:
  - GCM and dCRT perform similarly, consistent with asymptotic theory.
  - Lasso-based methods can have very inflated Type-I error in difficult settings.
  - Post-lasso-based dCRT and GCM typically outperform Maxway CRT.

Numerical simulations: Power

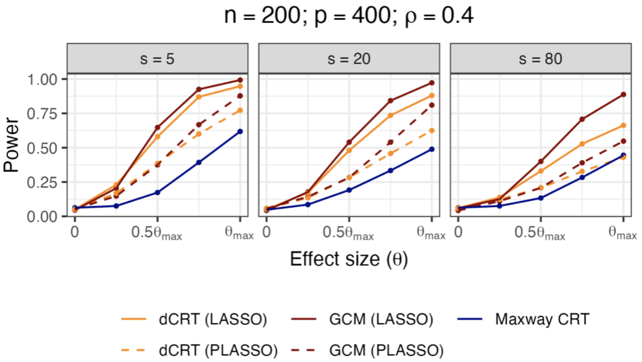


Numerical simulations: Power



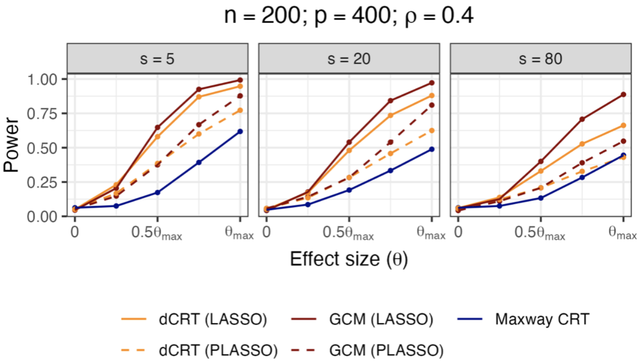
- Some takeaways:

Numerical simulations: Power



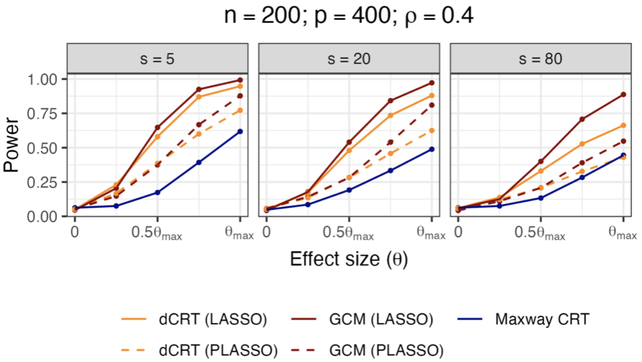
- Some takeaways:
  - GCM tends to outperform dCRT.

Numerical simulations: Power



- Some takeaways:
  - GCM tends to outperform dCRT.
  - Lasso outperforms post-lasso, suggesting bias-variance trade-off.

Numerical simulations: Power



- Some takeaways:
  - GCM tends to outperform dCRT.
  - Lasso outperforms post-lasso, suggesting bias-variance trade-off.
  - Maxway CRT has lowest power, due to data splitting.

References

Emmanuel Candes, Yingying Fan, Lucas Janson, and Jinchi Lv. Panning for gold:'model-x'knockoffs for high dimensional controlled variable selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(3):551–577, 2018.

Molei Liu, Eugene Katsevich, Lucas Janson, and Aaditya Ramdas. Fast and powerful conditional randomization testing via distillation. *Biometrika*, 109(2):277–293, 2022.

Ziang Niu, Abhinav Chakraborty, Oliver Dukes, and Eugene Katsevich. Reconciling model-x and doubly robust approaches to conditional independence testing. *arXiv preprint arXiv:2211.14698*, 2022.

Rajen D Shah and Jonas Peters. The hardness of conditional independence testing and the generalised covariance measure. *The Annals of Statistics*, 48(3):1514–1538, 2020.