UNC | GILLINGS SCHOOL OF GLOBAL PUBLIC HEALTH

Doubly robust estimation under a randomly censored covariate

Brian Richardson

Slides Available Online

Acknowledgements

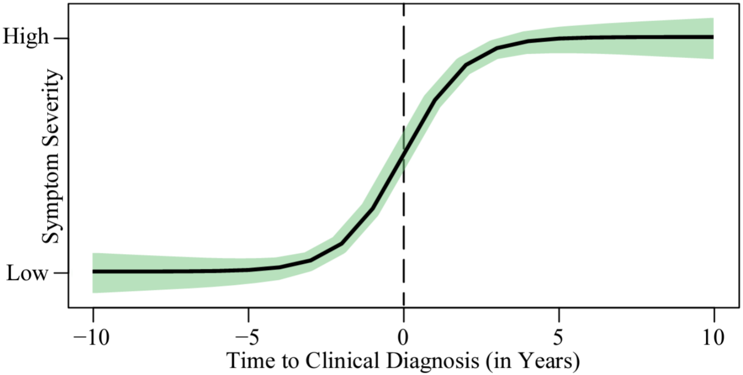
Seong-Ho Lee, PhD

Tanya Garcia, PhD

Yanyuan Ma, PhD

This research was supported by the National Institute of Environmental Health Sciences grant T32ES007018.

Huntington's Disease and Censored Covariates



Lotspeich et. al. (2024)

Censored Covariates: a Simple Example

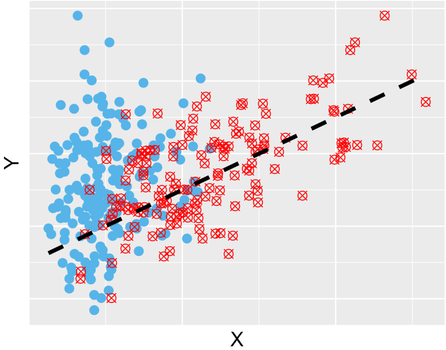
- Regression model: $\mathrm{E}(Y) = \beta_0 + \beta_1 X$

Censored Covariates: a Simple Example

- Regression model:
  $\mathrm{E}(Y) = \beta_0 + \beta_1 X$
- Estimate $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$ by solving:

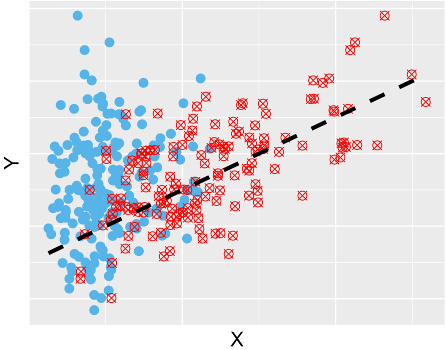
$$\sum_{i=1}^{n} \underbrace{(Y_i - \beta_0 - \beta_1 X_i)(1, X_i)^T}_{\text{score function: } \mathbf{S}^F_{\boldsymbol{\beta}}} = \mathbf{0}$$

Censored Covariates: a Simple Example



Problem: X is censored by a **random censoring time** C

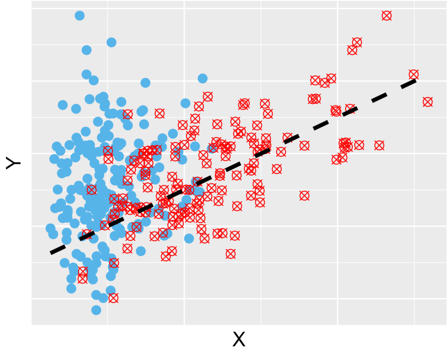
Censored Covariates: a Simple Example



Problem: X is censored by a **random censoring time** C

- $W = \min(X, C)$
- $\Delta = I(X \leq C)$

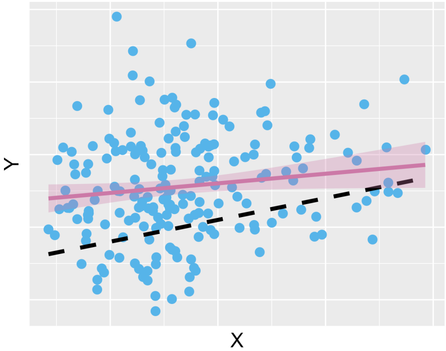
Censored Covariates: a Simple Example



Problem: X is censored by a **random censoring time** C

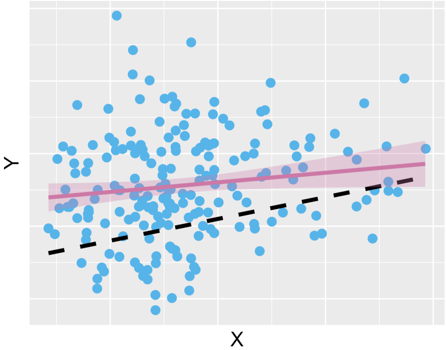
- $W = \min(X, C)$
- $\Delta = I(X \leq C)$
- assume: $C \perp\!\!\!\perp (X, Y)$

A Naive Approach



Naively treat W as X

A Naive Approach

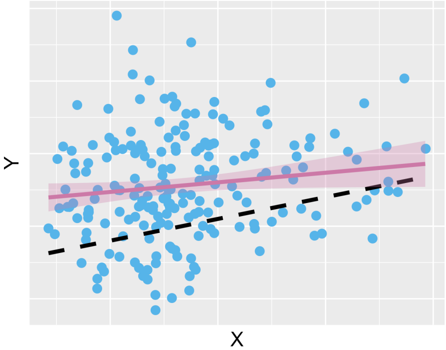


Naively treat W as X

- Estimate $\boldsymbol{\beta}$ by solving:

$$\sum_{i=1}^{n} \mathbf{S}_{\boldsymbol{\beta}}^{F}(Y_i, \boxed{W_i}, \boldsymbol{\beta}) = \mathbf{0}$$

A Naive Approach



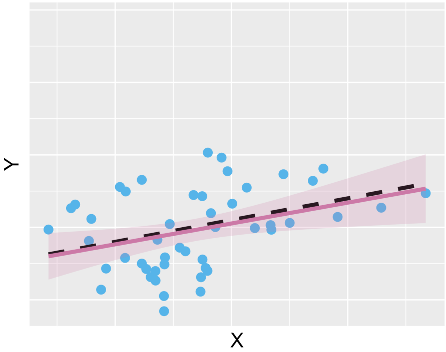
Naively treat W as X

- Estimate $\boldsymbol{\beta}$ by solving:

$$\sum_{i=1}^{n} \mathbf{S}^{F}_{\boldsymbol{\beta}}(Y_i, \boxed{W_i}, \boldsymbol{\beta}) = \mathbf{0}$$

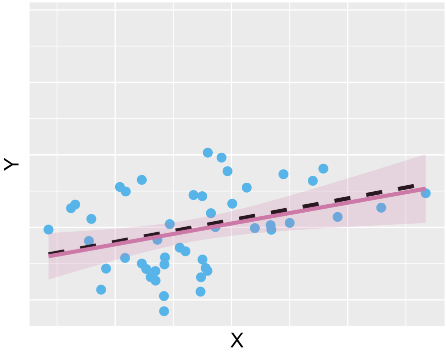
✗ Inconsistent estimator of $\boldsymbol{\beta}$

Complete Case Analysis



Only use observations where X is *uncensored*

Complete Case Analysis

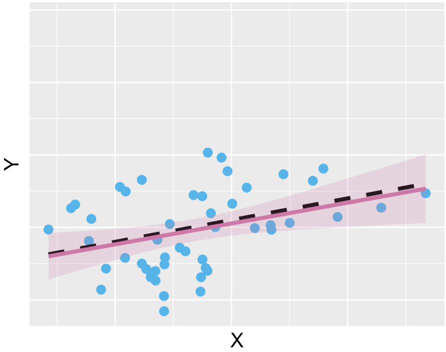


Only use observations where X is *uncensored*

- Estimate $\boldsymbol{\beta}$ by solving:

$$\sum_{i=1}^{n} \boxed{\Delta_i} \mathbf{S}^F_{\boldsymbol{\beta}}(Y_i, W_i, \boldsymbol{\beta}) = \mathbf{0}$$

Complete Case Analysis



Only use observations where X is *uncensored*

- Estimate $\boldsymbol{\beta}$ by solving:

$$\sum_{i=1}^{n} \boxed{\Delta_i} \mathbf{S}^F_{\boldsymbol{\beta}}(Y_i, W_i, \boldsymbol{\beta}) = \mathbf{0}$$

✓ Consistent

✗ Inefficient

Maximum Likelihood Estimation (MLE)

$$f_{Y,W,\Delta}(y, w, \delta, \boldsymbol{\beta}, \boxed{\boldsymbol{\alpha}}) \propto \underbrace{\{f_{Y|X}(y, w, \boldsymbol{\beta})\}^{\delta}}_{\text{uncensored}} \underbrace{\left\{\int_{w}^{\infty} f_{Y|X}(y, x, \boldsymbol{\beta}) \boxed{f_X(x, \boldsymbol{\alpha})} dx\right\}^{1-\delta}}_{\text{censored}}$$

Maximum Likelihood Estimation (MLE)

$$f_{Y,W,\Delta}(y, w, \delta, \boldsymbol{\beta}, \boxed{\boldsymbol{\alpha}}) \propto \underbrace{\{f_{Y|X}(y, w, \boldsymbol{\beta})\}^{\delta}}_{\text{uncensored}} \underbrace{\left\{\int_{w}^{\infty} f_{Y|X}(y, x, \boldsymbol{\beta}) \boxed{f_X(x, \boldsymbol{\alpha})} dx\right\}^{1-\delta}}_{\text{censored}}$$

$$\mathbf{S}_{\boldsymbol{\beta}}(y, w, \delta, \boldsymbol{\beta}) \equiv \partial \log f_{Y,W,\Delta}(y, w, \delta, \boldsymbol{\beta}, \alpha)/\partial \boldsymbol{\beta}$$

Maximum Likelihood Estimation (MLE)

$$f_{Y,W,\Delta}(y, w, \delta, \boldsymbol{\beta}, \boxed{\boldsymbol{\alpha}}) \propto \underbrace{\{f_{Y|X}(y, w, \boldsymbol{\beta})\}^{\delta}}_{\text{uncensored}} \underbrace{\left\{\int_{w}^{\infty} f_{Y|X}(y, x, \boldsymbol{\beta}) \boxed{f_X(x, \boldsymbol{\alpha})} dx\right\}^{1-\delta}}_{\text{censored}}$$

$$\mathbf{S}_{\boldsymbol{\beta}}(y, w, \delta, \boldsymbol{\beta}) \equiv \partial \log f_{Y,W,\Delta}(y, w, \delta, \boldsymbol{\beta}, \alpha)/\partial \boldsymbol{\beta}$$

$$\sum_{i=1}^{n} \mathbf{S}_{\boldsymbol{\beta}}(Y_i, W_i, \Delta_i, \boldsymbol{\beta}) = \mathbf{0}$$

Maximum Likelihood Estimation (MLE)

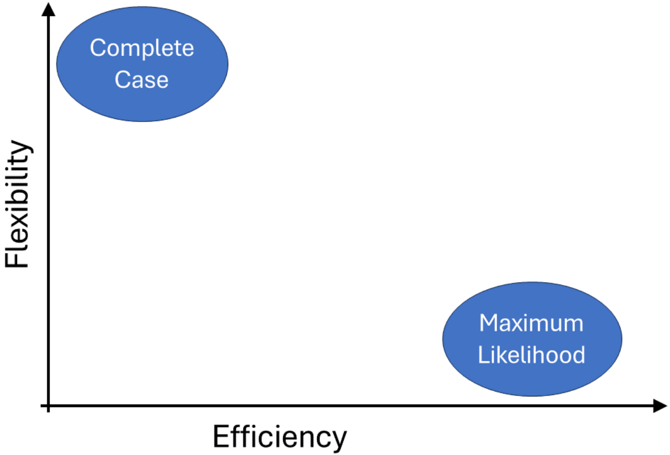
Pros

- ✓ consistent estimator of β
- ✓ fully efficient

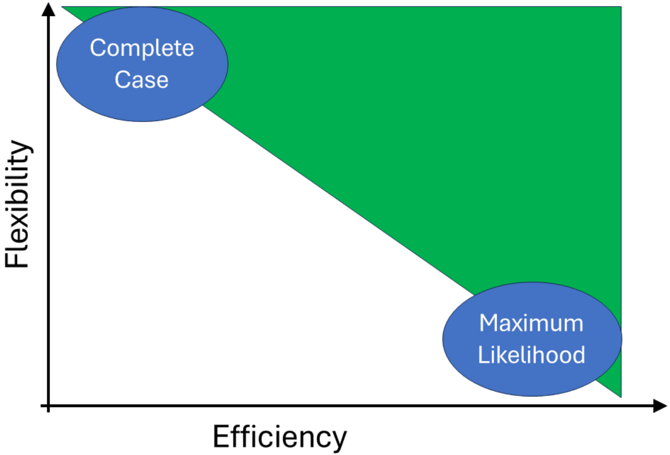
Cons

- ✗ requires model for **nuisance parameter** f_X
- ✗ inconsistent estimator when model for f_X is incorrect

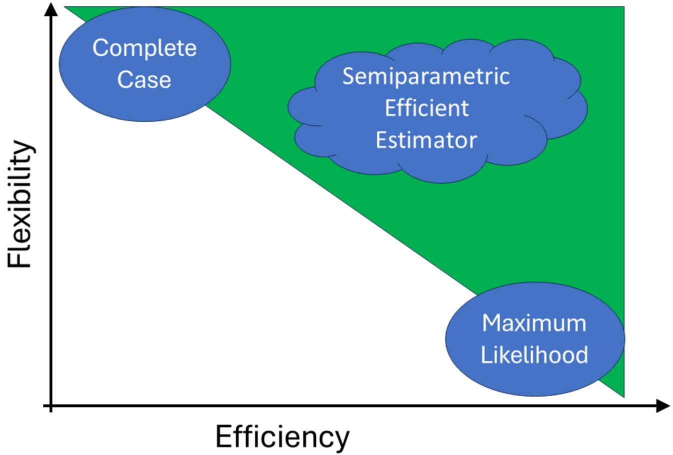
Existing Methods



Existing Opportunity



A New Approach



The Semiparametric Recipe

- **goal**: to find the estimating function resulting in a semiparametric efficient estimator

The Semiparametric Recipe

- **goal**: to find the estimating function resulting in a semiparametric efficient estimator

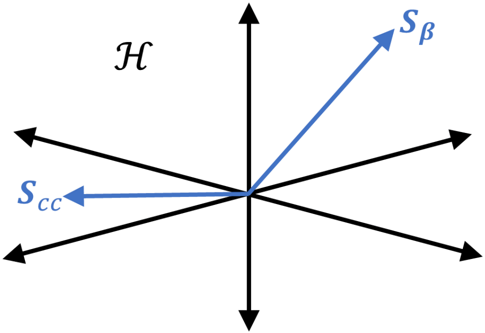
- **semiparametric**: infinite dimensional nuisance parameters f_X, f_C

The Semiparametric Recipe

- **goal**: to find the estimating function resulting in a semiparametric efficient estimator
- **semiparametric**: infinite dimensional nuisance parameters f_X, f_C
- Geometric approach from Tsiatis (2006)

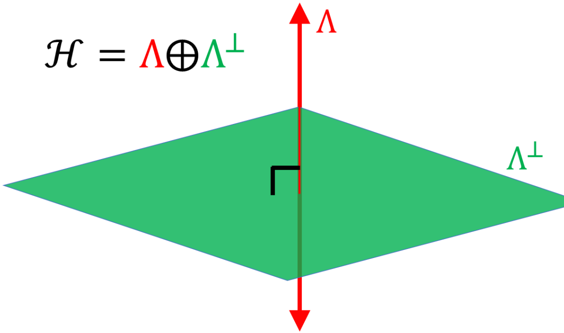
The Semiparametric Recipe

- **Hilbert space** of estimating functions
- **covariance inner product** $\langle \boldsymbol{h}, \boldsymbol{g} \rangle = \mathrm{E}(\boldsymbol{h}^T \boldsymbol{g})$

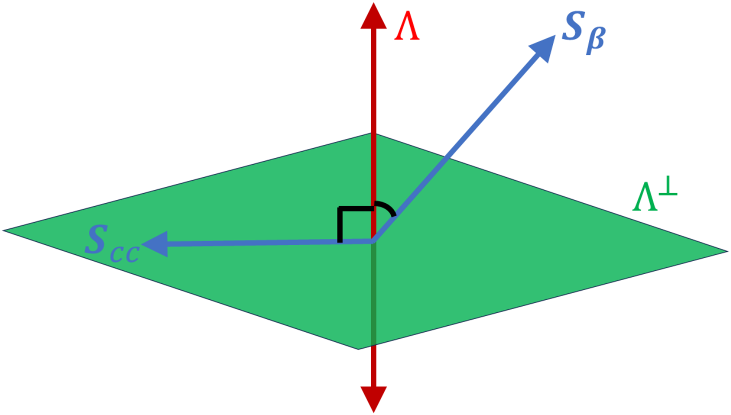


The Semiparametric Recipe

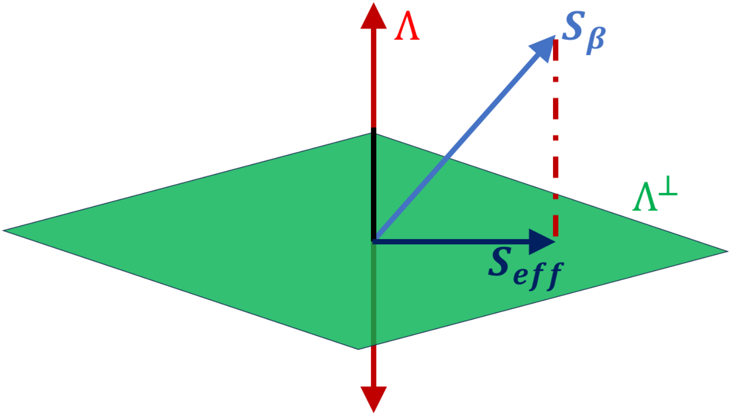
- construct Λ using **nuisance scores**
- orthogonal complement $\Lambda^\perp$



The Semiparametric Recipe



The Semiparametric Recipe



Properties of the Proposed Estimator

The **semiparametric efficient estimator** $\widehat{\boldsymbol{\beta}}_{\text{eff}}$ is the solution to

$$\sum_{i=1}^{n} \mathbf{S}_{\text{eff}}^{F}(Y_i, W_i, \Delta_i, \boldsymbol{\beta}) = \mathbf{0}$$

Properties of the Proposed Estimator

The **semiparametric efficient estimator** $\widehat{\boldsymbol{\beta}}_{\text{eff}}$ is the solution to

$$\sum_{i=1}^{n} \mathbf{S}_{\text{eff}}^{F}(Y_i, W_i, \Delta_i, \boldsymbol{\beta}) = \mathbf{0}$$

✓ consistent and asymptotically normal

Properties of the Proposed Estimator

The **semiparametric efficient estimator** $\widehat{\boldsymbol{\beta}}_{\text{eff}}$ is the solution to

$$\sum_{i=1}^{n} \mathbf{S}_{\text{eff}}^{F}(Y_i, W_i, \Delta_i, \boldsymbol{\beta}) = \mathbf{0}$$

- ✓ consistent and asymptotically normal
- ✓ **Doubly Robust**: $\widehat{\boldsymbol{\beta}}_{\text{eff}}$ is consistent if at least one of f_X, f_C is correctly specified

Properties of the Proposed Estimator

The **semiparametric efficient estimator** $\widehat{\boldsymbol{\beta}}_{\text{eff}}$ is the solution to

$$\sum_{i=1}^{n} \mathbf{S}_{\text{eff}}^{F}(Y_i, W_i, \Delta_i, \boldsymbol{\beta}) = \mathbf{0}$$

- ✓ consistent and asymptotically normal
- ✓ **Doubly Robust**: $\widehat{\boldsymbol{\beta}}_{\text{eff}}$ is consistent if at least one of f_X, f_C is correctly specified
- ✓ **Locally Efficiency**: If f_X, f_C are *both* correctly specified, then $\widehat{\boldsymbol{\beta}}_{\text{eff}}$ achieves the **semiparametric efficiency bound**

Simulation Setup

- $Y|X \sim N(\beta_0 + \beta_1 X, \sigma^2)$

Simulation Setup

- $Y|X \sim N(\beta_0 + \beta_1 X, \sigma^2)$
- large sample size $n = 10,000$

Simulation Setup

- $Y|X \sim N(\beta_0 + \beta_1 X, \sigma^2)$
- large sample size $n = 10,000$
- high censoring rate $q = \mathrm{P}(X > C) = 0.8$

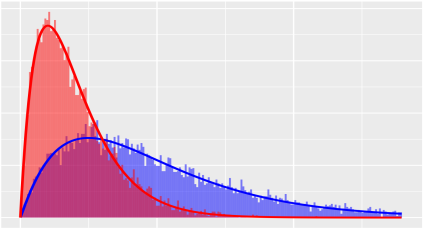
Simulation Setup

- $Y|X \sim N(\beta_0 + \beta_1 X, \sigma^2)$
- large sample size $n = 10,000$
- high censoring rate $q = \mathrm{P}(X > C) = 0.8$
- $X, C \sim$ gamma distributions

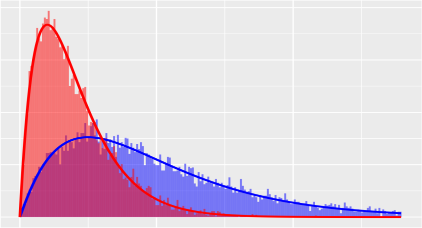
Simulation Setup

- $Y|X \sim N(\beta_0 + \beta_1 X, \sigma^2)$
- large sample size $n = 10,000$
- high censoring rate $q = \mathrm{P}(X > C) = 0.8$
- X, $C \sim$ gamma distributions
- X, C possibly **misspecified** as exponential

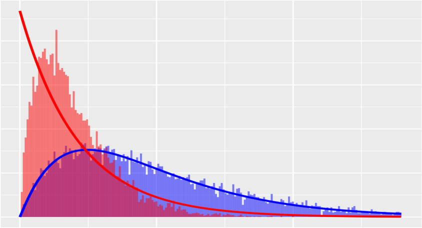
X, C correct



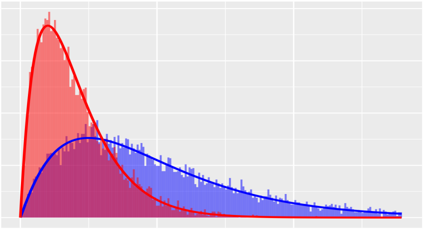
X, C correct



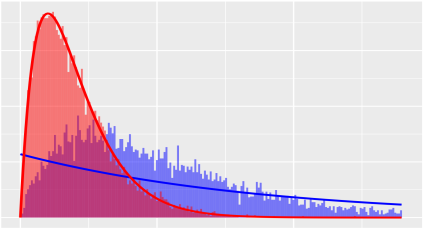
X correct, C incorrect



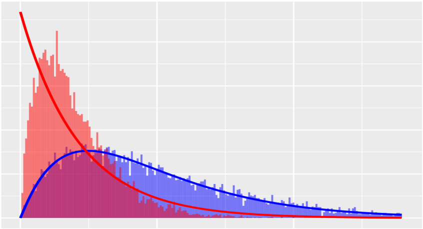
X, C correct



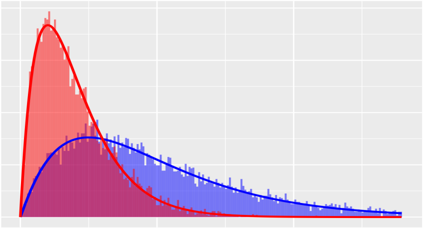
X incorrect, C correct



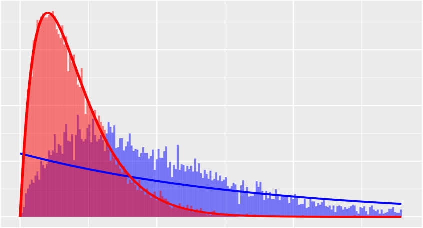
X correct, C incorrect



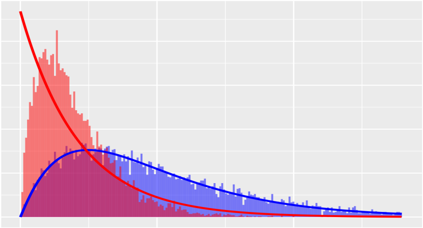
X, C correct



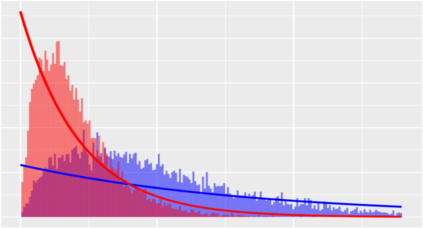
X incorrect, C correct



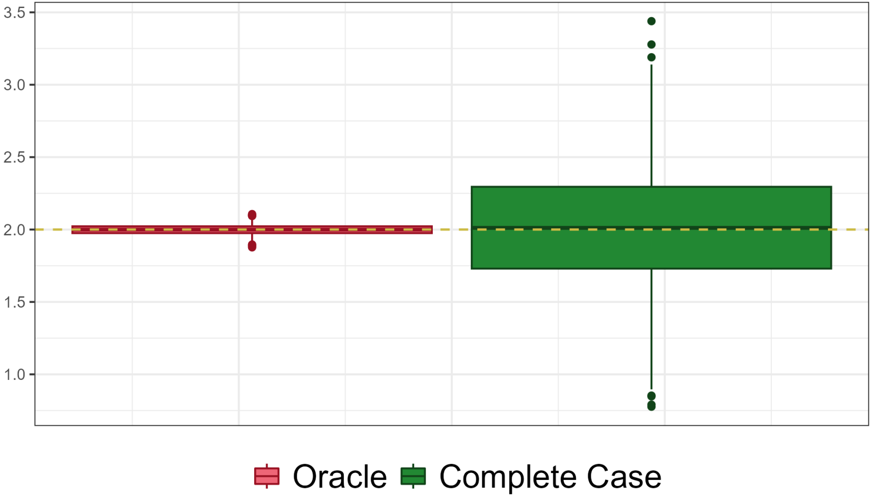
X correct, C incorrect



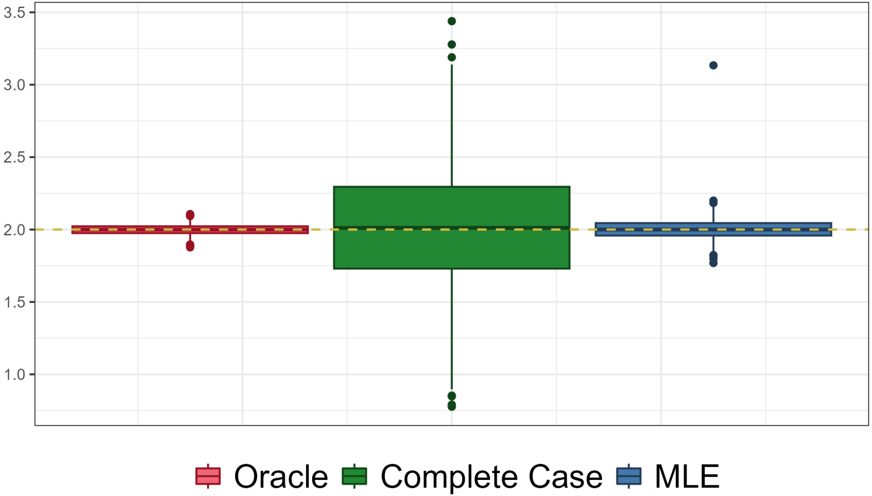
X, C incorrect



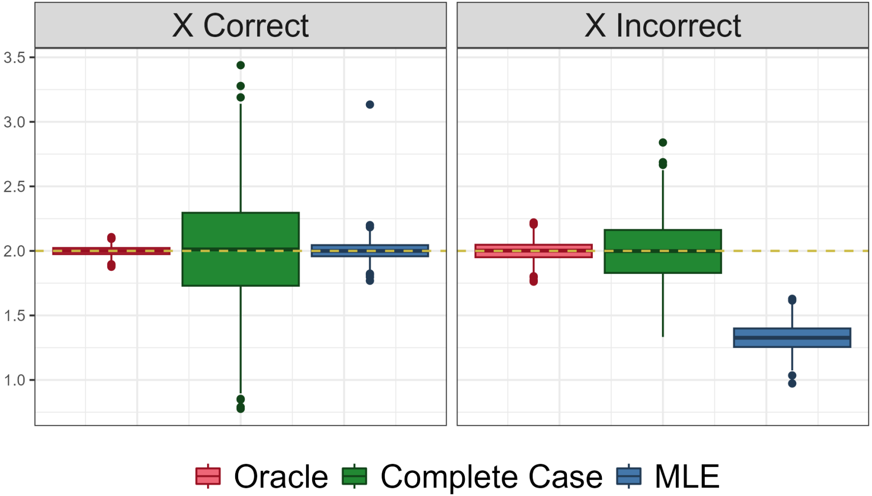
Simulation Results



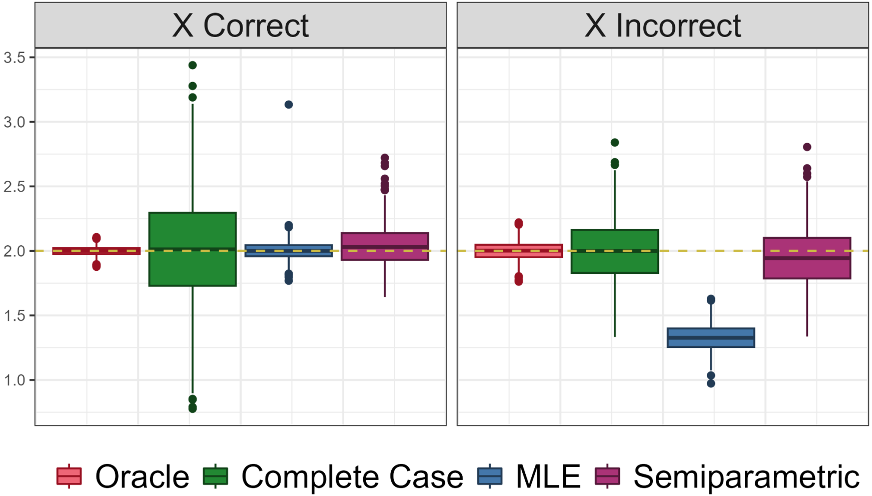
Simulation Results



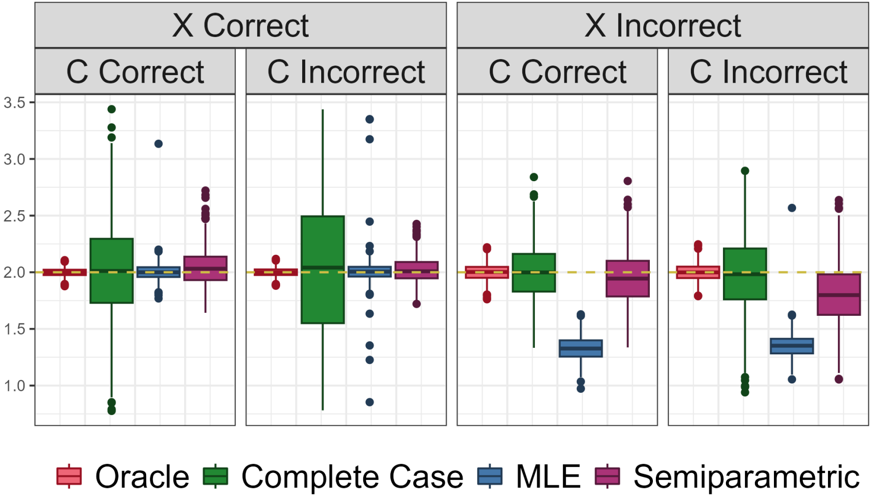
Simulation Results



Simulation Results



Simulation Results



Generalizations

The methods presented here extend to:

- Nonlinear $\mathrm{E}(Y|X) = m(X, \boldsymbol{\beta})$

Generalizations

The methods presented here extend to:

- Nonlinear $\mathrm{E}(Y|X) = m(X, \boldsymbol{\beta})$
- Additional uncensored covariates $\mathbf{Z}$
  - $\mathrm{E}(Y|X, \mathbf{Z}) = m(X, \mathbf{Z}, \boldsymbol{\beta})$
  - Nuisance distributions become $f_{X|\mathbf{Z}}, f_{C|\mathbf{Z}}, f_{\mathbf{Z}}$

SPARCC: Semiparametric Censored Covariate Estimation

R package available at https://github.com/brian-d-richardson/sparcc

Thank you! Any questions?

Brian Richardson
🖂: brichson@ad.unc.edu