Optimal Nuisance Function Tuning for Estimating a Doubly Robust Functional under Proportional Asymptotics

Sean McGrath[1], Debarghya Mukherjee[2], Rajarshi Mukherjee[3], Zixiao Jolene Wang[3]

[1]Yale University, [2]Boston University, [3]Harvard University

NeurIPS 2025

Problem Setup

- **Motivating Question:** How do we tune estimators of nuisance parameters for optimal downstream estimation of the parameter of interest?
- **Data:** We observe n i.i.d. copies of (A, Y, X), where $A, Y \in \mathbb{R}$, $X \in \mathbb{R}^p$
- **Parameter of Interest:**

$$\begin{aligned} \theta_0 &= \mathbb{E}\left[\mathrm{Cov}(A, Y|X)\right] \\ &= \mathbb{E}\left[\{A - \mathbb{E}[A|X]\}\{Y - \mathbb{E}[Y|X]\}\right] \end{aligned}$$

- **Model:** Coordinates of X are i.i.d. subgaussian and

$$\begin{pmatrix} A \\ Y \end{pmatrix} |X \sim \text{Normal}\left(\begin{pmatrix} X^\top \alpha_0 \\ X^\top \beta_0 \end{pmatrix}, \begin{pmatrix} 1 & \theta_0 \\ \theta_0 & 1 \end{pmatrix}\right)$$

- **Asymptotic Regime:** $p/n \to c \in (0, \infty)$

Estimators

- **Estimators of Nuisance Parameters:**
  - $\hat{\alpha}(\lambda_1)$: Ridge regression estimator of α_0 with tuning parameter λ_1
  - $\hat{\beta}(\lambda_2)$: Ridge regression estimator of β_0 with tuning parameter λ_2

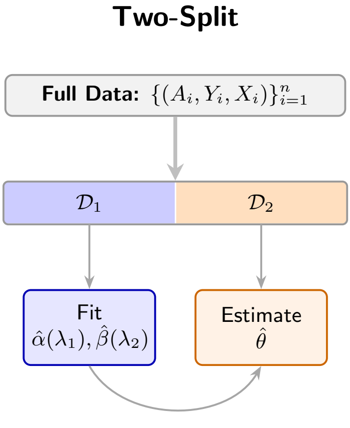
- **(Naive) Estimators of θ_0:**

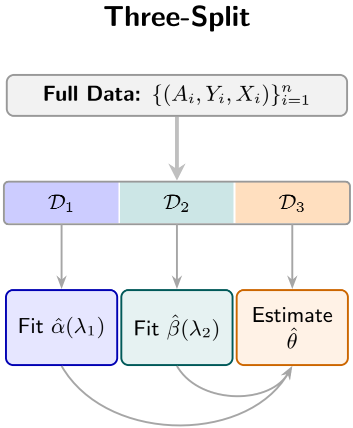
$$\hat{\theta}^{\mathrm{INT}} = \frac{1}{n}\sum_{i=1}^{n} A_i Y_i - \hat{\alpha}(\lambda_1)^\top \hat{\beta}(\lambda_2) \qquad \text{(Integral-Based)}$$

$$\hat{\theta}^{\mathrm{NR}} = \frac{1}{n}\sum_{i=1}^{n} Y_i (A_i - X_i^\top \hat{\alpha}(\lambda_1)) \qquad \text{(Newey-Robins)}$$

$$\hat{\theta}^{\mathrm{DR}} = \frac{1}{n}\sum_{i=1}^{n} (A_i - X_i^\top \hat{\alpha}(\lambda_1))(Y_i - X_i^\top \hat{\beta}(\lambda_2)) \qquad \text{(Doubly Robust)}$$

Sample Splitting Strategies





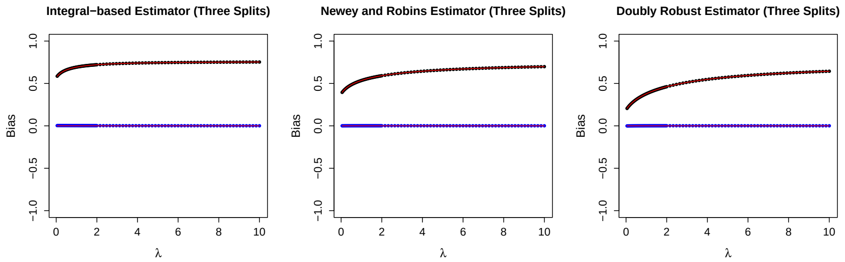
Main Results

- **Debiasing Step:** Each of the three estimators has $\Theta(1)$ bias in each sample splitting strategy
  - Need to consider (further) debiased estimators!
- **Theorems 3.1–3.3:** For each estimator in each sample splitting strategy, we construct a debiased version of the estimator, $\hat{\theta}^{\mathrm{db}}$, such that

$$\sqrt{n}(\hat{\theta}^{\mathrm{db}} - \theta_0) = O_p(1)$$

- **Theorem 3.4:** For each debiased estimator $\hat{\theta}^{\mathrm{db}}$, we characterize the limiting variance $\lim_{n\to\infty} n \cdot \mathrm{Var}(\hat{\theta}^{\mathrm{db}})$
- **Resulting Estimation Strategy:** Select (λ_1, λ_2) to minimize the limiting variance of the debiased estimator

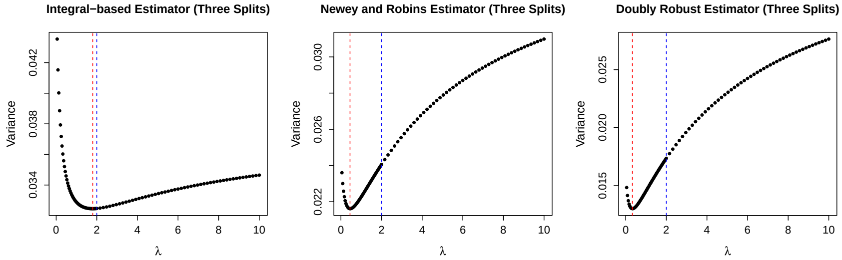
Numerical Results: Bias of Naive and Debiased Estimators



- **Conclusion:** Naive estimators (black dots) show clear bias, whereas debiased estimators (blue dots) indeed show negligible bias

We consider the setting where $\lambda_1 = \lambda_2 =: \lambda$

Numerical Results: Variance of Debiased Estimators



- **Conclusion:** Optimal choices of λ for estimating θ_0 (dotted red line) differ from those that optimally estimate α_0, β_0 (dotted blue line)

We consider the setting where $\lambda_1 = \lambda_2 =: \lambda$