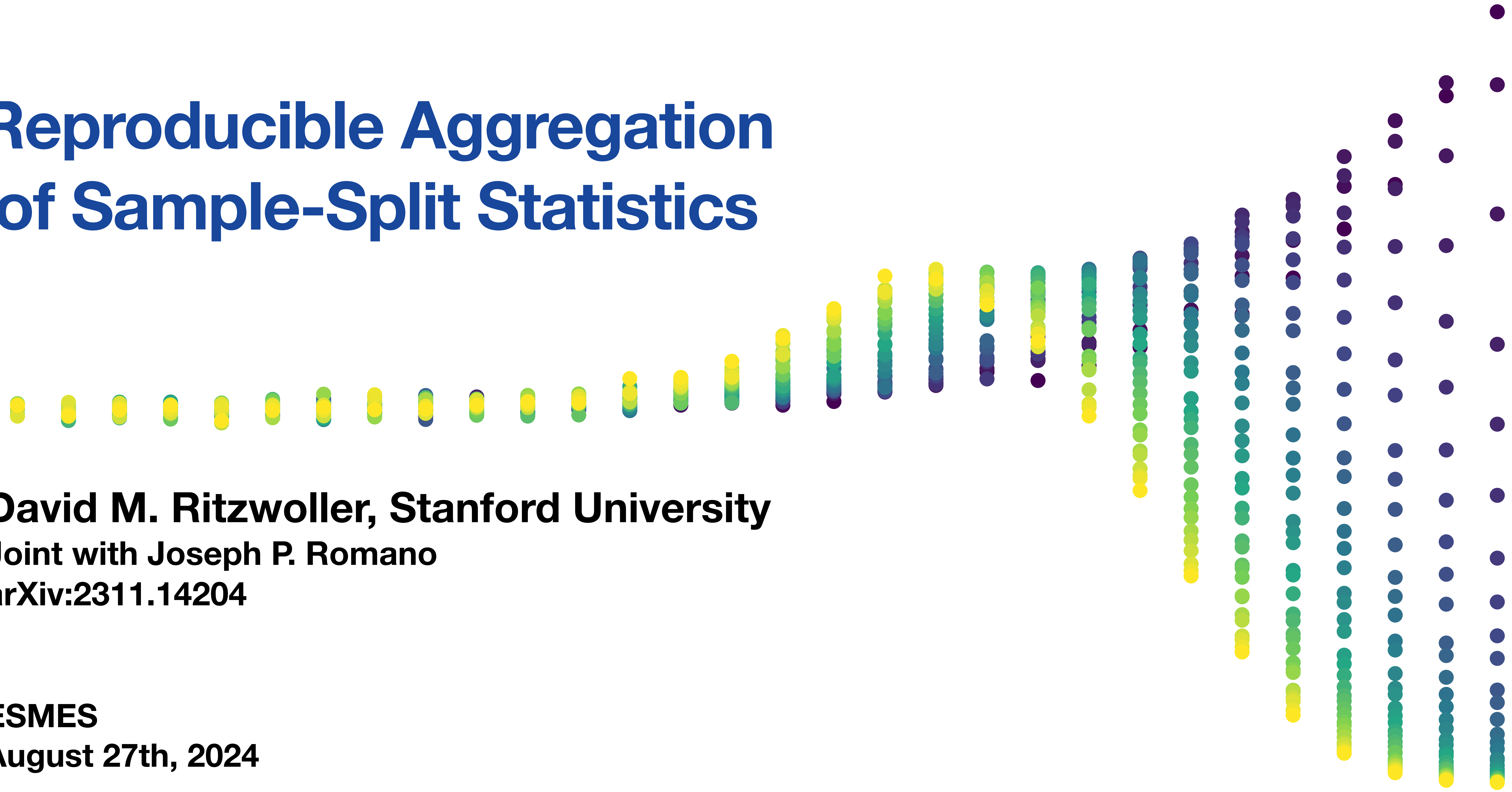
Reproducible Aggregation of Sample-Split Statistics



David M. Ritzwoller, Stanford University
Joint with Joseph P. Romano
arXiv:2311.14204

ESMES
August 27th, 2024

Banerjee et. al (2015) evaluate a poverty alleviation program

W_i - Assignment to treatment

X_i - Pretreatment covariates

Y_i - Consumption three years after implementation

Collect the data $D_i = (Y_i, W_i, X_i)$ into $D = (D_i)_{i=1}^n$

Banerjee et. al (2015) evaluate a poverty alleviation program

W_i - Assignment to treatment

X_i - Pretreatment covariates

Y_i - Consumption three years after implementation

Collect the data $D_i = (Y_i, W_i, X_i)$ into $D = (D_i)_{i=1}^n$

Augmented Inverse Propensity Score Weighting (Robbins et. al, 1994)

- Split the sample into D_{s} and $D_{\tilde{\mathsf{s}}}$, where $(\mathsf{s}, \tilde{\mathsf{s}})$ partition $[n] = \{1, \ldots, n\}$
- Compute

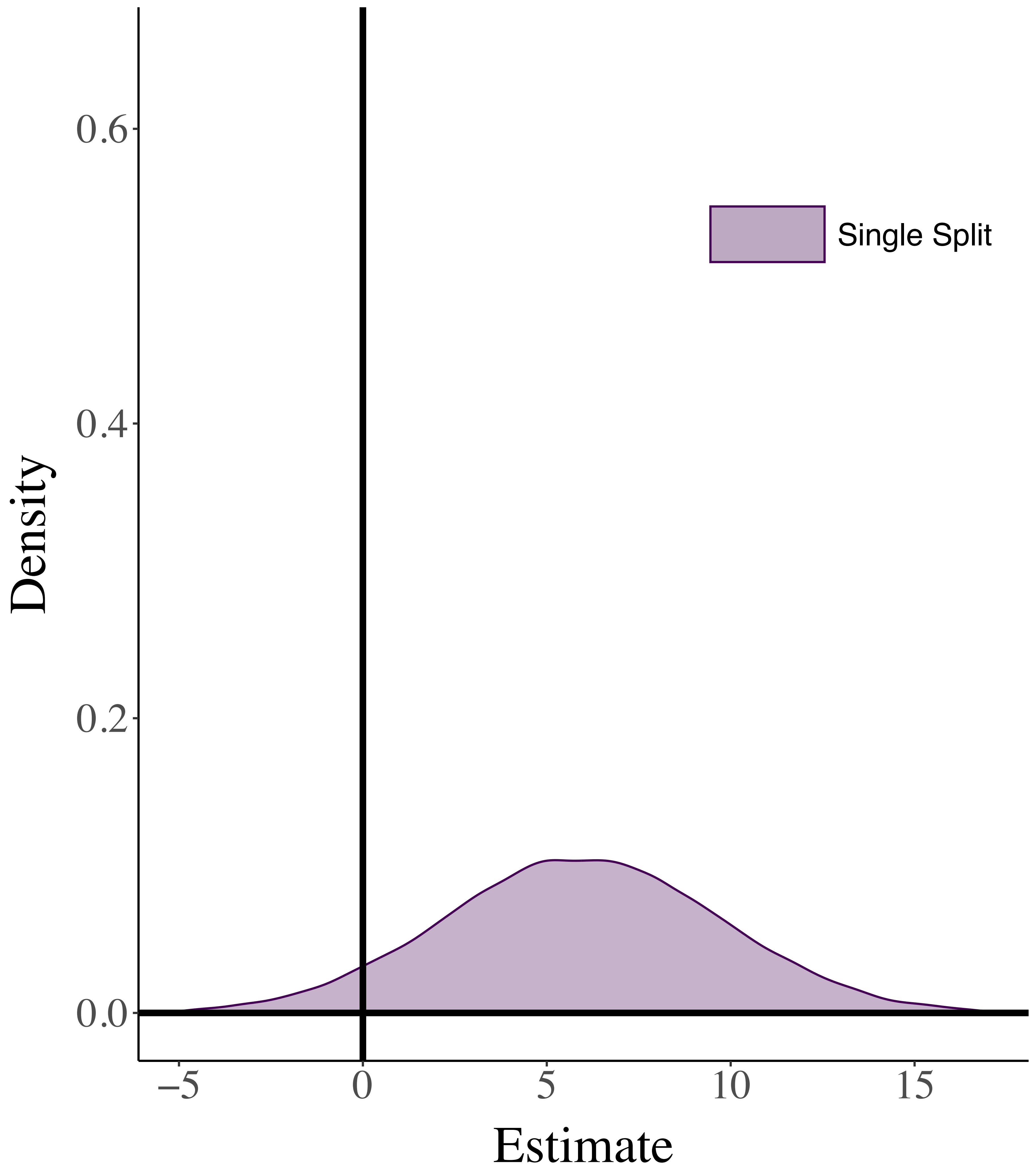
$$T(\mathsf{s}, D) = \frac{1}{|\mathsf{s}|} \sum_{i \in \mathsf{s}} \psi(D_i, \hat{\eta}(D_{\tilde{\mathsf{s}}}))$$

$$\psi(D_i, \hat{\eta}(D_{\tilde{\mathsf{s}}})) = \hat{\mu}_1(X_i) - \hat{\mu}_0(X_i) + \frac{W_i(Y_i - \hat{\mu}_1(X_i))}{\hat{\pi}(X_i)} - \frac{(1 - W_i)(Y_i - \hat{\mu}_0(X_i))}{1 - \hat{\pi}(X_i)}$$

where $\hat{\eta}(D_{\tilde{\mathsf{s}}}) = (\hat{\pi}, \hat{\mu}_w)$ collects nuisance estimates constructed with $D_{\tilde{\mathsf{s}}}$

Let s be a random subset of $[n]$ of size $n/2$

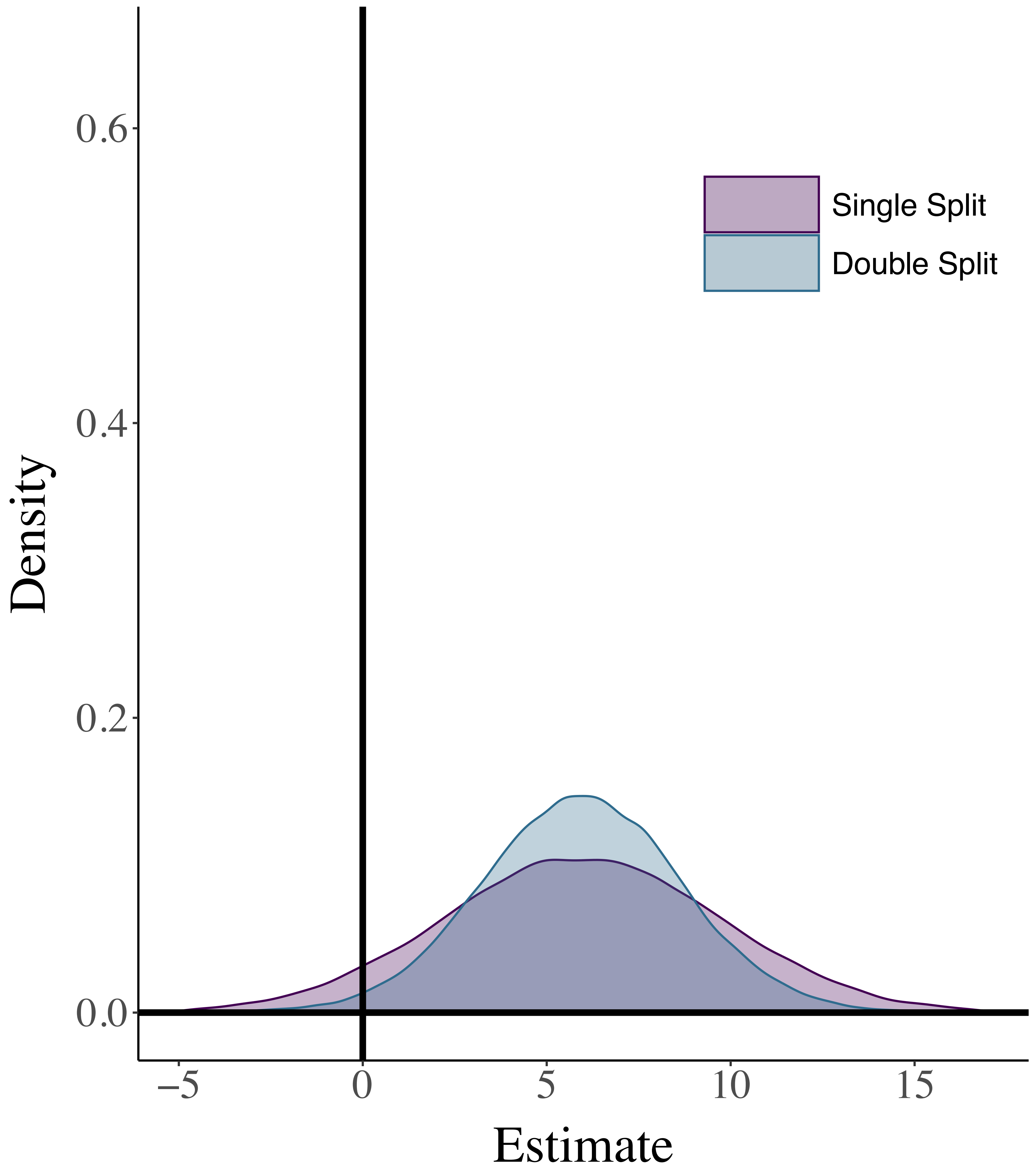
Single-Split: $T(\mathsf{s}, D)$



Let s and s$'$ be random subsets of $[n]$ of size $n/2$

Single-Split: $T(\mathsf{s}, D)$

Double-Split: $\frac{1}{2}\left(T(\mathsf{s}, D) + T(\mathsf{s}', D)\right)$

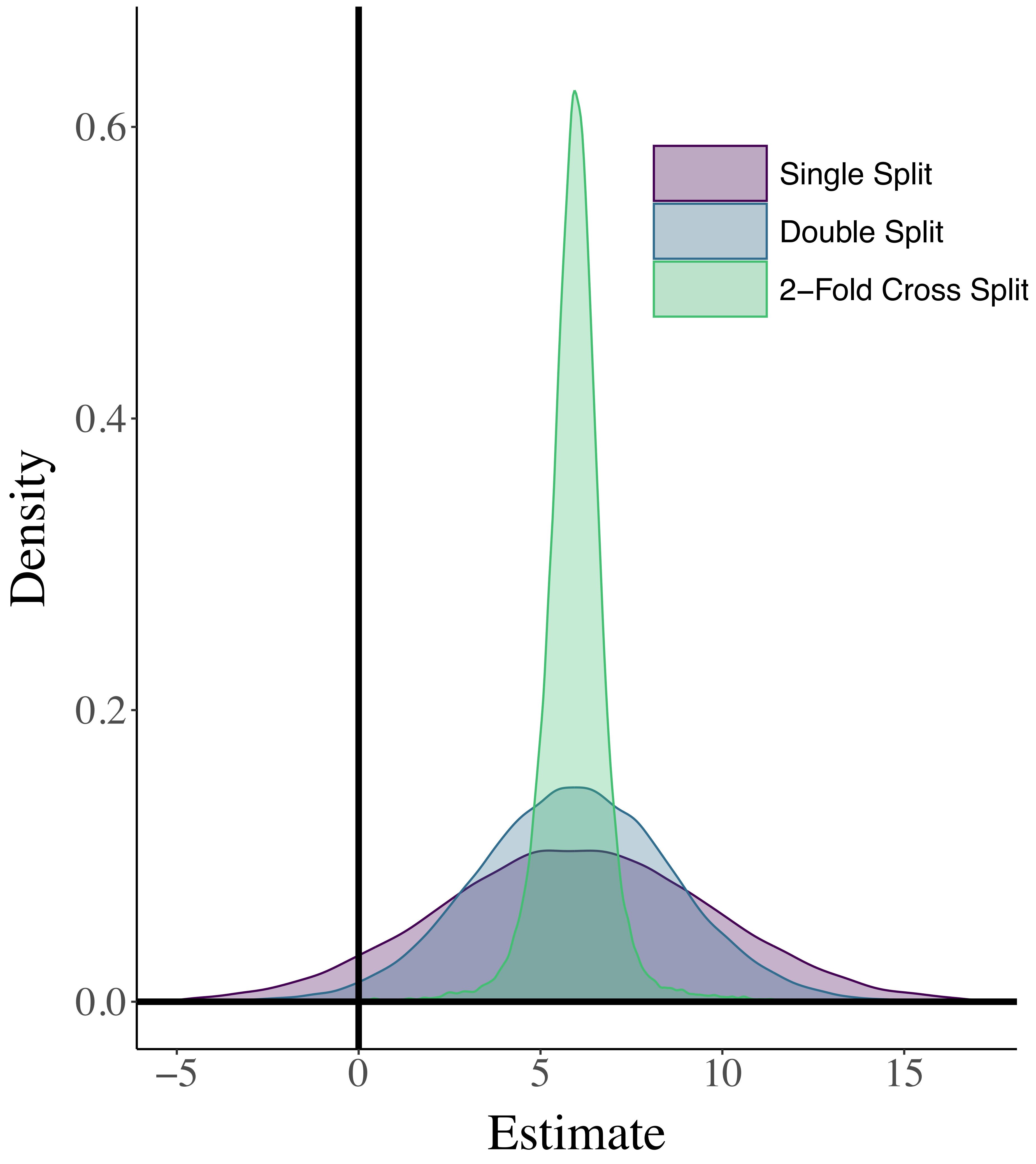


Let s and s' be random subsets of $[n]$ of size $n/2$

Single-Split: $T(\mathsf{s}, D)$

Double-Split: $\frac{1}{2}\left(T(\mathsf{s}, D) + T(\mathsf{s}', D)\right)$

Two-Fold Cross-Split: $\frac{1}{2}\left(T(\mathsf{s}, D) + T(\tilde{\mathsf{s}}, D)\right)$



A moderate increase in the number of sample-splits does not do away with the problem

Consider the k-fold cross-split estimator

$$a(\mathsf{r}_k, D) = \frac{1}{k} \sum_{j=1}^{k} T(\mathsf{s}_j, D) \ ,$$

where $\mathsf{r}_k = (\mathsf{s}_i)_{i=1}^{k}$ is some k-fold partition of $[n]$

The associated critical value is given by

$$\mathrm{CV}_\alpha = \frac{z_{1-\alpha}}{n} \left(\sum_{j=1}^{k} \sum_{i \in \mathsf{s}_j} (\psi(D_i, \hat{\eta}(D_{\tilde{\mathsf{s}}_j})) - a(\mathsf{r}_k, D))^2 \right)^{1/2}$$

Take $k = 10$

A moderate increase in the number of sample-splits does not do away with the problem

Consider the k-fold cross-split estimator

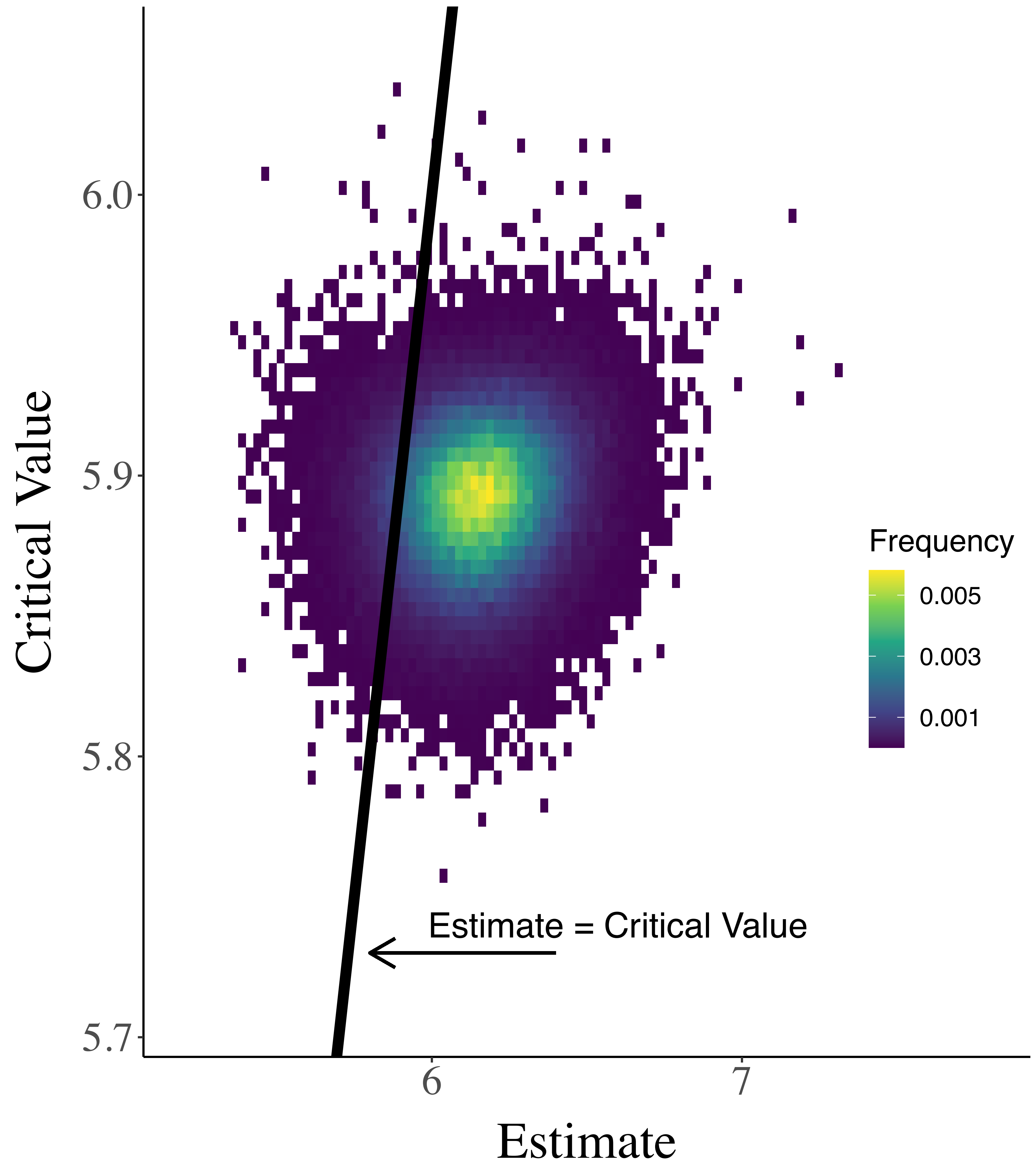
$$a(\mathsf{r}_k, D) = \frac{1}{k}\sum_{j=1}^{k} T(\mathsf{s}_j, D)\ ,$$

where $\mathsf{r}_k = (\mathsf{s}_i)_{i=1}^k$ is some k-fold partition of $[n]$

The associated critical value is given by

$$\mathsf{CV}_\alpha = \frac{z_{1-\alpha}}{n}\left(\sum_{j=1}^{k}\sum_{i\in\mathsf{s}_j}\left(\psi(D_i, \hat{\eta}(D_{\tilde{\mathsf{s}}_j})) - a(\mathsf{r}_k, D)\right)^2\right)^{1/2}$$

Take $k = 10$



If anything, the problem appears more severe in an application to risk estimation

Here, we consider the sample-split risk estimate

$$T(\mathsf{s}, D) = \frac{1}{\sum_{i \in \mathsf{s}} \mathbb{I}\{W_i = 1\}} \sum_{i \in \mathsf{s}} \mathbb{I}\{W_i = 1\}(Y_i - \hat{\beta}_1(\lambda)^\top X_i)^2$$

where $\hat{\beta}_1(\lambda)$ is the Lasso coefficient estimated on the sample $D_{\tilde{s}}$ with the penalization parameter λ

Cross-Validation: Aggregate with k-fold cross-splitting in the same way. Select λ minimizing estimated risk.

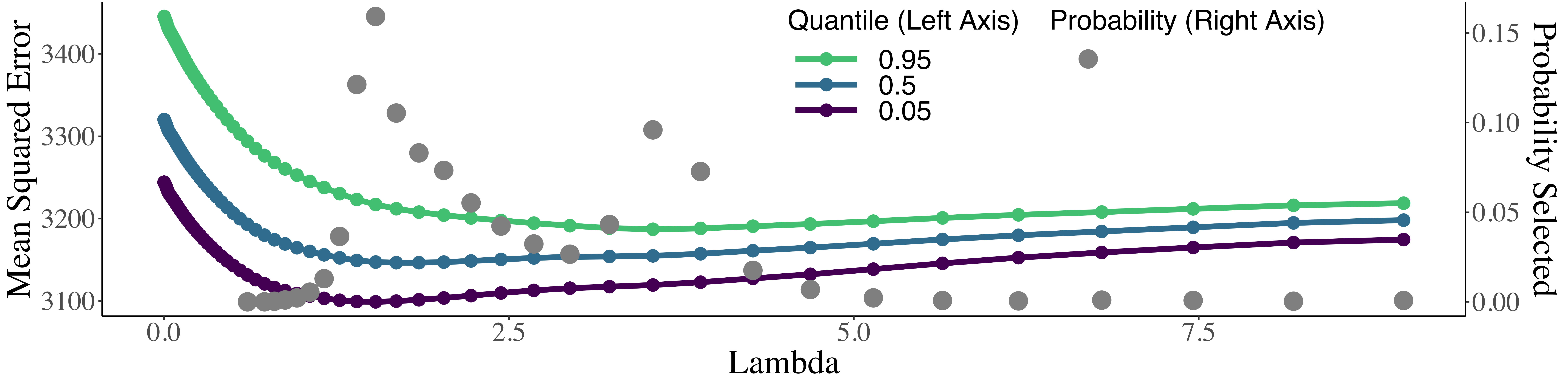
If anything, the problem appears more severe in an application to risk estimation

Here, we consider the sample-split risk estimate

$$T(\mathsf{s}, D) = \frac{1}{\sum_{i \in \mathsf{s}} \mathbb{I}\{W_i = 1\}} \sum_{i \in \mathsf{s}} \mathbb{I}\{W_i = 1\}(Y_i - \hat{\beta}_1(\lambda)^\top X_i)^2$$

where $\hat{\beta}_1(\lambda)$ is the Lasso coefficient estimated on the sample $D_{\tilde{\mathsf{s}}}$ with the penalization parameter λ

Cross-Validation: Aggregate with k-fold cross-splitting in the same way. Select λ minimizing estimated risk.



This Paper

1. Methodology

Algorithm: Propose a simple procedure for sequentially aggregating sample-split statistics

Input: User chooses a bound and an error rate

Objective: Probability that residual randomness is smaller than the bound is less than error rate

2. Theory

1. Establish validity of procedure, in particular asymptotic sense
2. Concentration result, characterizing difference between cross-splitting and independent splitting
3. Berry-Esseen bound, illustrating trade-off between computational efficiency and accuracy

Related Literature

1. Sequential Statistics
   (i) Classical Methods: Anscombe (1952), Chow and Robbins (1965)
2. Cross-Validation and Cross Splitting
   (i) Inference for Generalization Error: Bayle, Bayle, Janson, and Mackey (2020), Austern and Zhou (2020)
   (ii) Algorithmic Stability: Kale, Kumar, and Vassilvitskii (2011), Kumar, Lokshtanov, Vassilvitskii, and Vattani (2013), Chen, Syrgkanis, and Austern (2022), Ritzwoller and Syrgkanis (2024)
   (iii) Additional Applications: DiCiccio, DiCiccio, and Romano (2020), Wasserman, Ramdas, and Balakrishnan (2020), Ramdas and Manole (2023)
3. Stein's Method
   (i) Exchangeable Pairs: Chatterjee (2005, 2007), Paulin, Mackey, and Tropp (2013, 2016)

Outline

1. Proposal and Generic Validity
2. Non-Asymptotic Theory
   (i) Concentration and Normal Approximation
   (ii) Reproducibility
3. Performance

Notation

- The set $\mathscr{S}_{n,b}$ contains all subsets of $[n]$ of size b
- The set $\mathscr{R}_{n,b}$ is the collection partitions of $[n]$ into sets of size b

Notation

- The set $\mathscr{S}_{n,b}$ contains all subsets of $[n]$ of size b
- The set $\mathscr{R}_{n,b}$ is the collection partitions of $[n]$ into sets of size b

Problem

Consider a sample $D = (D_i)_{i=1}^{n}$. We are interested in reporting a real-valued sample-split statistic

$$T(\mathsf{s}, D) = \Psi(D_{\mathsf{s}}, \hat{\eta}(D_{\tilde{\mathsf{s}}}))$$

where $\hat{\eta}(\,\cdot\,)$ is an estimator of an unknown nuisance parameter and s is in $\mathscr{S}_{n,b}$

We study aggregate statistics of the form

$$a(\mathsf{R}_{g,k}, D) = \frac{1}{g}\frac{1}{k}\sum_{i=1}^{g}\sum_{j=1}^{k} T(\mathsf{s}_{i,j}, D)$$

where $\mathsf{R}_{g,k} = (\mathsf{r}_i)_{i=1}^{g}$ collects g elements of $\mathscr{R}_{n,k,b}$ and we write $\mathsf{r}_i = (\mathsf{s}_{i,j})_{j=1}^{k}$

Reproducibility

Our task is to choose the number of collections of mutually exclusive splits g to ensure that the residual variability in the aggregate statistic $a(\mathsf{R}_{g,k}, D)$ is small

Definition: Reproducibility

Suppose that the integers $\hat{g}$ and $\hat{g}'$ and the collections $\mathsf{R}_{\hat{g},k}$ and $\mathsf{R}'_{\hat{g}',k}$ are independent and identically distributed, conditional on the data D

We say that $a(\mathsf{R}_{\hat{g},k}, D)$ is (ξ, β)-reproducible if

$$P\left\{ \,|\, a(\mathsf{R}_{\hat{g},k}, D) - a(\mathsf{R}'_{\hat{g}',k}, D) \,|\, \leq \xi \mid D \right\} \geq 1 - \beta$$

almost surely

Anscombe-Chow-Robbins Aggregation

We propose a sequential method for constructing a reproducible statistic $a(\mathsf{R}_{\hat{g},k}, D)$

Define the variance estimator

$$\hat{v}(\mathsf{R}_{g,k}, D) = \frac{1}{g(g-1)} \sum_{j=1}^{g} (a(\mathsf{r}_j, D) - a(\mathsf{R}_{g,k}, D))^2$$

where we recall that $\mathsf{R}_{g,k} = (\mathsf{r}_j)_{j=1}^{g}$

Algorithm: Anscombe-Chow-Robbins Aggregation

Let $\hat{g}$ be the smallest value of g greater than or equal to g_{init} such that the condition

$$\hat{v}(\mathsf{R}_{g,k}, D) \leq \mathsf{cv}_{\xi,\beta} = \frac{1}{2}\left(\frac{\xi}{z_{1-\beta/2}}\right)^2$$

is satisfied. Return $a(\mathsf{R}_{\hat{g},k}, D)$.

Generic Validity

Theorem: Generic Validity

If the collections $\mathsf{R}_{\hat{g},k}$ and $\mathsf{R}'_{\hat{g}',k}$ are independently computed with the Anscombe-Chow-Robbins procedure, then

$$P\left\{ \,|\, a(\mathsf{R}_{\hat{g},k}, D) - a(\mathsf{R}'_{\hat{g}',k}, D) \,|\, \geq \xi \mid D \right\} \to \beta$$

as $\xi \to 0$

Proof Sketch: Define $v_{g,k}(D) = \mathrm{Var}(a(\mathsf{R}_{g,k}, D) \mid D)$. We show that $\hat{g}/g^\star \to 1$ a.s. as $\xi \to 0$, where

$$g^\star = \min_g \left\{ \mathrm{Var}(a(\mathsf{R}_{g,k}, D) \mid D) \leq \frac{1}{2}\left(\frac{\xi}{z_{1-\beta/2}}\right)^2 \right\}, \quad \text{i.e.,} \quad \xi \approx z_{1-\beta/2}\sqrt{2 \cdot v_{g^\star,k}(D))} \;.$$

A central limit theorem holds for $a(\mathsf{R}_{g^\star,k}, D)$. The discrepancy $|\, a(\mathsf{R}_{\hat{g},k}, D) - a(\mathsf{R}_{g^\star,k}, D) \,|$ can be bounded. ■

Taking Stock

To this point, we have made no assumptions. But have we gained a real statistical understanding of the problem? At least two questions arise:

Taking Stock

To this point, we have made no assumptions. But have we gained a real statistical understanding of the problem? At least two questions arise:

1. Have we said anything about cross-splitting?
   - On inspection, the proof only uses the independence of the g collections r_i in $\mathrm{R}_{g,k}$

Taking Stock

To this point, we have made no assumptions. But have we gained a real statistical understanding of the problem? At least two questions arise:

1. Have we said anything about cross-splitting?

2. How should we interpret the approximation with $\xi \to 0$?
   - This asymptotic is somewhat nebulous, or at least, unfamiliar
   - How does the associated approximation depend on parameters like k or $g^\star$?

Taking Stock

To this point, we have made <u>no</u> assumptions. But have we gained a real statistical understanding of the problem? At least two questions arise:

1. Have we said anything about cross-splitting?
2. How should we interpret the approximation with $\xi \to 0$?
    - This asymptotic is somewhat nebulous, or at least, unfamiliar
    - How does the associated approximation depend on parameters like k or $g^\star$?

Objective: Quantify the accuracy of the nominal error rate β in a way that accounts for and compares the concentration in $a(\mathsf{R}_{g,k}, D)$ with both g and k

Outline

1. Proposal and Generic Validity
2. Non-Asymptotic Theory
   (i) Concentration and Normal Approximation
   (ii) Reproducibility
3. Performance

Outline

1. Proposal and Generic Validity
2. Non-Asymptotic Theory
   (i) Concentration and Normal Approximation
   (ii) Reproducibility
3. Performance

Symmetry and Linearity

We impose two simplifying restrictions on the statistic of interest

Assumption: Symmetry and Determinism

For all sets $\mathsf{s} \subseteq [n]$ and data D, the statistic $T(\mathsf{s}, D)$ is deterministic and invariant to permutations of the data with indices in s and of the data with indices in $\tilde{\mathsf{s}}$, respectively.

Intention: Restrict randomness under consideration to randomness induced by sample-splitting

Symmetry and Linearity

We impose two simplifying restrictions on the statistic of interest

Assumption: Symmetry and Determinism

For all sets $\mathsf{s} \subseteq [n]$ and data D, the statistic $T(\mathsf{s}, D)$ is deterministic and invariant to permutations of the data with indices in s and of the data with indices in $\tilde{\mathsf{s}}$, respectively.

Intention: Restrict randomness under consideration to randomness induced by sample-splitting

Assumption: Linearity

For all sets $\mathsf{s} \subseteq [n]$ and data D, the statistic $T(\mathsf{s}, D)$ can be written

$$T(\mathsf{s}, D) = \Psi(D_{\mathsf{s}}, \hat{\eta}(D_{\tilde{\mathsf{s}}})) = \frac{1}{|\mathsf{s}|} \sum_{i \in \mathsf{s}} \psi(D_i, \hat{\eta}(D_{\tilde{\mathsf{s}}}))$$

for some function $\psi(\,\cdot\,,\cdot\,)$

Note: Easily relaxed to bounded differences, component-wise Lipschitz, etc.

Stability

Our results are expressed in terms of the following objects

- Let D' denote an independent copy of the data D
- For each set q $\subseteq [n]$, let $\tilde{D}^{(q)}$ be constructed by swapping D_i with D_i' for each i in q

Definition: Stability

Fix a set s $\in \mathcal{S}_{n,b}$. Let q be a randomly selected subset of $\tilde{\mathrm{s}}$ of cardinality q. We refer to the quantity

$$\sigma^{(r,q)} = \mathbb{E}\left[\,|\, \psi(D_i, \hat{\eta}(D_{\tilde{\mathrm{s}}})) - \psi(D_i, \hat{\eta}(\tilde{D}^{(\mathrm{q})}_{\tilde{\mathrm{s}}})) \,|^r \right]$$

as the (r, q)th-order training stability.

Stability

At times, we restrict attention to statistics satisfying the following bound

Definition: Stability

A statistic $T(\mathsf{s}, D)$ is stable if

$$\sigma^{(r,q)} \lesssim \left(\frac{q}{n-b}\right)^r$$

for all positive even integers r

- Holds (and is tight) if $\hat{\eta}$ is an empirical risk minimizer of a strictly convex loss
- Widely studied in the statistical learning literature, e.g.,
  - Subsampled regression (Chen, Syrgkanis, and Austern, 2022, Ritzwoller and Syrgkanis, 2024)
  - Stochastic gradient descent (Hardt, Recht, and Singer, 2016)

Concentration

The following result characterizes the concentration of $a(\mathsf{R}_{g,k}, D)$ around $\bar{a}(D) = \mathbb{E}[a(\mathsf{R}_{g,k}, D) \mid D]$

Theorem: Large Deviation Bound

Let $\varphi = b/n$. Under the stated assumptions, the inequality

$$P\left\{ |a(\mathsf{R}_{g,k}, D) - \bar{a}(D)| \lesssim (1-\varphi)\sqrt{\frac{\sigma^{(2,b-1)}}{g}\frac{\log(\varepsilon^{-1})}{\delta}} \mid D \right\} \geq 1 - \varepsilon$$

holds with probability greater than $1 - \delta$ as D varies

Challenge: Handling dependence in summands of $a(\mathsf{R}_{g,k}, D)$ across cross-splits

Approach: Use a coupling argument to construct an exchangeable pair, apply Stein's method (Chatterjee, 2005, 2007)

Suppose that $T(\mathsf{s}, D)$ is stable. The rate reduces to

$$\sqrt{\frac{\sigma^{(2,b-1)}}{g}} \underset{\text{Stability}}{\lesssim} \frac{1}{\sqrt{g}}\frac{1}{k}$$

Normal Approximation

The Anscombe-Chow-Robbins procedure depends on a normal approximation

Theorem: Berry-Esseen Bound

Define the normalized statistic

$$U(\mathsf{R}_{g,k}, D) = \frac{a(\mathsf{R}_{g,k}, D) - \bar{a}(D)}{(v_{g,k}(D))^{1/2}}$$

Under the stated assumptions, if Z is standard normal, then the inequality

$$\sup_{z \in \mathbb{R}} \left(P\{U(\mathsf{R}_{g,k}, D) \leq z \mid D\} - P\{Z \leq z\} \right) \lesssim \frac{1}{\delta} \frac{(1-\varphi)^3}{\sqrt{g}} \left(\frac{(\sigma^{(4,b-1)})^{1/2}}{v_{1,k}(D)} \right)^{3/2}$$

holds with probability greater than $1 - \delta$ as D varies

Normal Approximation

Consider the error in the normal approximation

$$\frac{(1-\varphi)^3}{\sqrt{g}}\left(\frac{(\sigma_{\text{train}}^{(4,b-1)})^{1/2}}{v_{1,k}(D)}\right)^{3/2}$$

By stability and an upper bound on $v_{1,k}(D)$ derived from the concentration inequality, we can show that

$$\frac{(1-\varphi)^3}{\sqrt{g}}\left(\frac{(\sigma_{\text{train}}^{(4,b-1)})^{1/2}}{v_{1,k}(D)}\right)^{3/2} \underset{\text{Stability}}{\gtrsim} \frac{1}{\sqrt{g}}$$

Normal Approximation

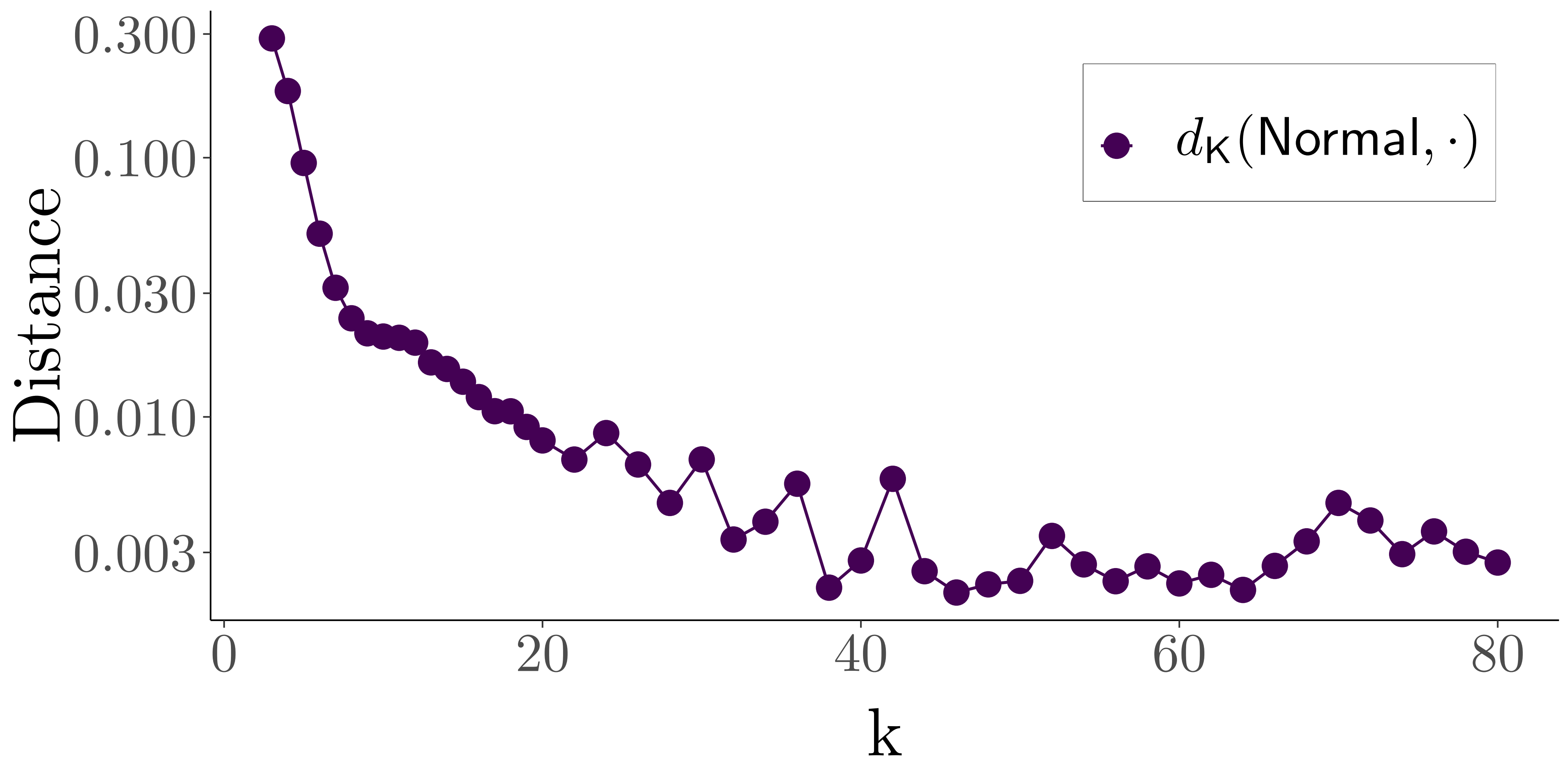
Consider the error in the normal approximation

$$\frac{(1-\varphi)^3}{\sqrt{g}}\left(\frac{(\sigma_{\text{train}}^{(4,b-1)})^{1/2}}{v_{1,k}(D)}\right)^{3/2}$$

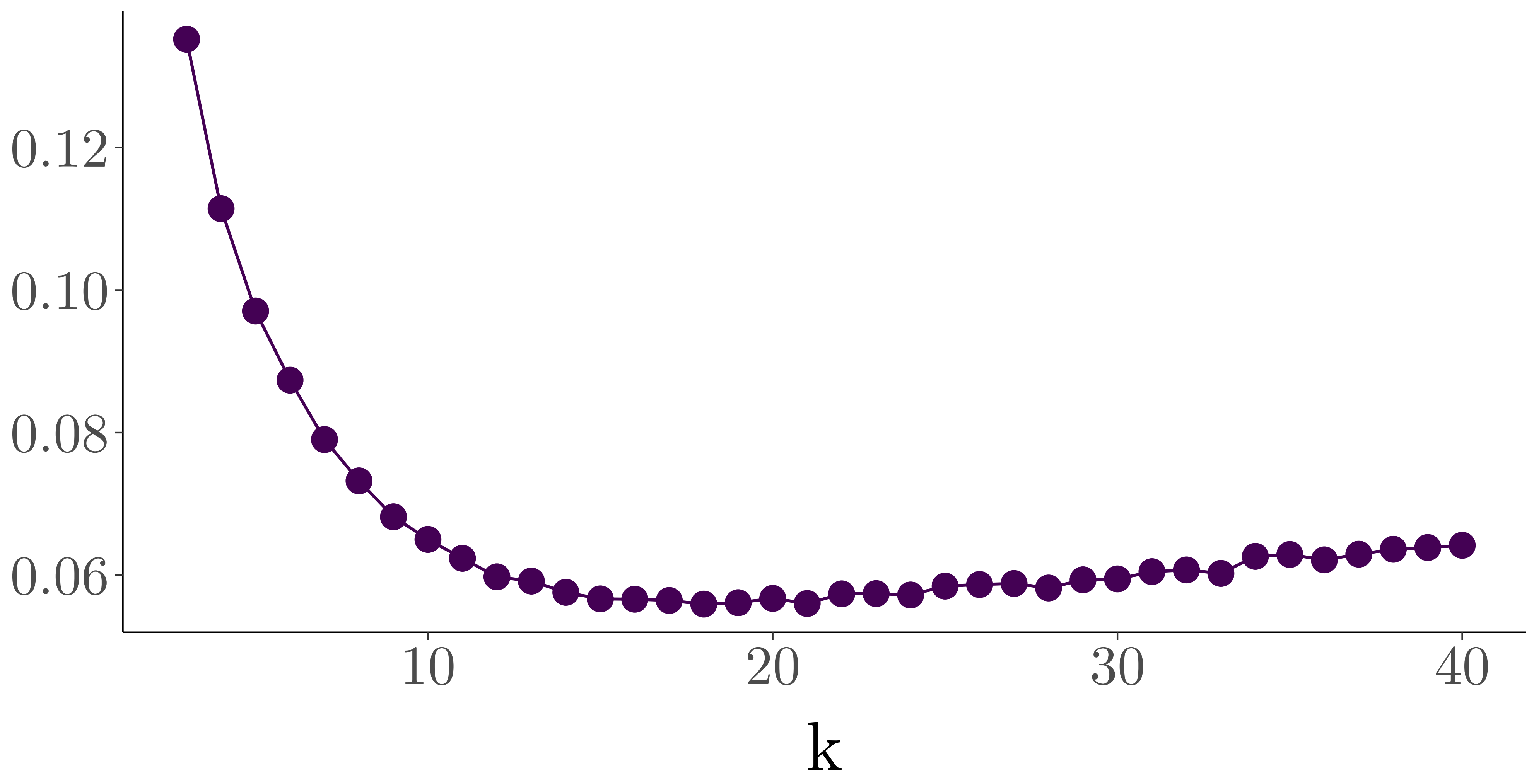
By stability and an upper bound on $v_{1,k}(D)$ derived from the concentration inequality, we can show that

$$\frac{(1-\varphi)^3}{\sqrt{g}}\left(\frac{(\sigma_{\text{train}}^{(4,b-1)})^{1/2}}{v_{1,k}(D)}\right)^{3/2} \underset{\text{Stability}}{\gtrsim} \frac{1}{\sqrt{g}}$$

Panel A: Treatment Effect Estimation



Panel B: Risk Estimation



Outline

1. Proposal and Generic Validity
2. Non-Asymptotic Theory
   (i) Concentration and Normal Approximation
   (ii) Reproducibility
3. Performance

Reproducibility

We're now equipped to re-consider the Anscombe-Chow-Robbins procedure

Theorem: Berry-Esseen Bound

If the collections $\mathsf{R}_{\hat{g},k}$ and $\mathsf{R}'_{\hat{g}',k}$ are independently computed with the Anscombe-Chow-Robbins procedure, then

$$\left| P\left\{ \, | a(\mathsf{R}_{\hat{g},k}, D) - a(\mathsf{R}'_{\hat{g}',k}, D) | \leq \xi \mid D \right\} - (1-\beta) \right| \leq \mathsf{A} + \mathsf{B}$$

holds with probability greater than $1-\delta$ as D varies, where

$$\mathsf{A} = \frac{1}{\delta} \frac{\xi}{z_{1-\beta/2}} \frac{(1-\varphi)^3 (\sigma^{(4,b-1)})^{3/4}}{(v_{1,k}(D))^2} \quad \text{and} \quad \mathsf{B} = \left(\frac{1}{\delta^{3/2}} \frac{\xi}{z_{1-\beta/2}} \frac{(1-\varphi)^4 \sigma^{(2,b-1)} (\sigma^{(4,b-1)})^{1/2}}{(v_{1,k}(D))^{5/2}} \right)^{1/2}$$

- A results from a normal approximation to $a(\mathsf{R}_{g^\star,k}, D) - a(\mathsf{R}'_{g^\star,k}, D)$
- B results from the difference $a(\mathsf{R}_{\hat{g},k}, D) - a(\mathsf{R}_{g^\star,k}, D)$
- The dependence on ξ is optimal (Landers and Rogge, 1976)

Computation and Accuracy

Computation

- The oracle stopping time $g^\star$ is proportional to

$$g^\star \underset{\text{Stability}}{\approx} \frac{2}{k^2}\left(\frac{z_{1-\beta/2}}{\xi}\right)$$

- The total number of splits used by the oracle procedure $m^\star = kg^\star$ is proportional to k^{-1}

Computation and Accuracy

Computation

- The oracle stopping time $g^\star$ is proportional to

$$g^\star \underset{\text{Stability}}{\approx} \frac{2}{k^2}\left(\frac{z_{1-\beta/2}}{\xi}\right)$$

- The total number of splits used by the oracle procedure $m^\star = kg^\star$ is proportional to k^{-1}

Accuracy

- The leading term in the accuracy of the nominal error rate is proportional to

$$\left(\frac{\xi}{z_{1-\beta/2}} \frac{(1-\varphi)^4 \sigma_{\text{train}}^{(4,b-1)} (\sigma_{\text{train}}^{(4,b-1)})^{1/2}}{(v_{1,k}(D))^{5/2}}\right)^{1/2} \underset{\text{Stability}}{\approx} \left(\frac{\xi k}{z_{1-\beta/2}}\right)^{1/2} \approx (g^\star)^{-1/4}$$

Computation and Accuracy

Computation

- The oracle stopping time $g^\star$ is proportional to

$$g^\star \underset{\text{Stability}}{\approx} \frac{2}{k^2}\left(\frac{z_{1-\beta/2}}{\xi}\right)$$

- The total number of splits used by the oracle procedure $m^\star = kg^\star$ is proportional to k^{-1}

Accuracy

- The leading term in the accuracy of the nominal error rate is proportional to

$$\left(\frac{\xi}{z_{1-\beta/2}} \frac{(1-\varphi)^4 \sigma_{\text{train}}^{(4,b-1)} (\sigma_{\text{train}}^{(4,b-1)})^{1/2}}{(v_{1,k}(D))^{5/2}}\right)^{1/2} \underset{\text{Stability}}{\approx} \left(\frac{\xi k}{z_{1-\beta/2}}\right)^{1/2} \approx (g^\star)^{-1/4}$$

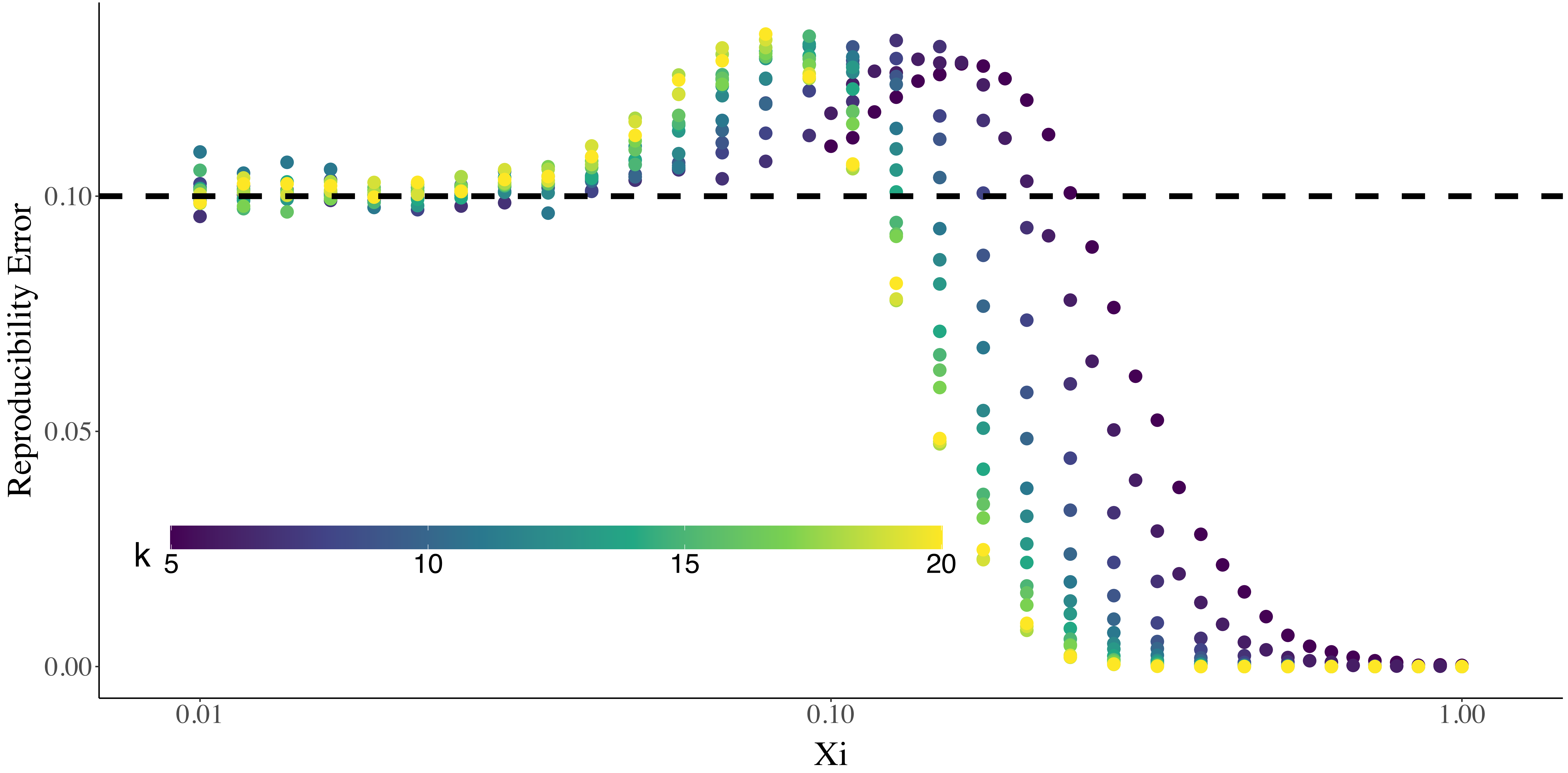
Upshot

- There is a fundamental tradeoff between computation and accuracy
- The rate $(g^\star)^{-1/4}$ is slower than for non-sequential problems (i.e., usually coverage error is order $n^{-1/2}$)

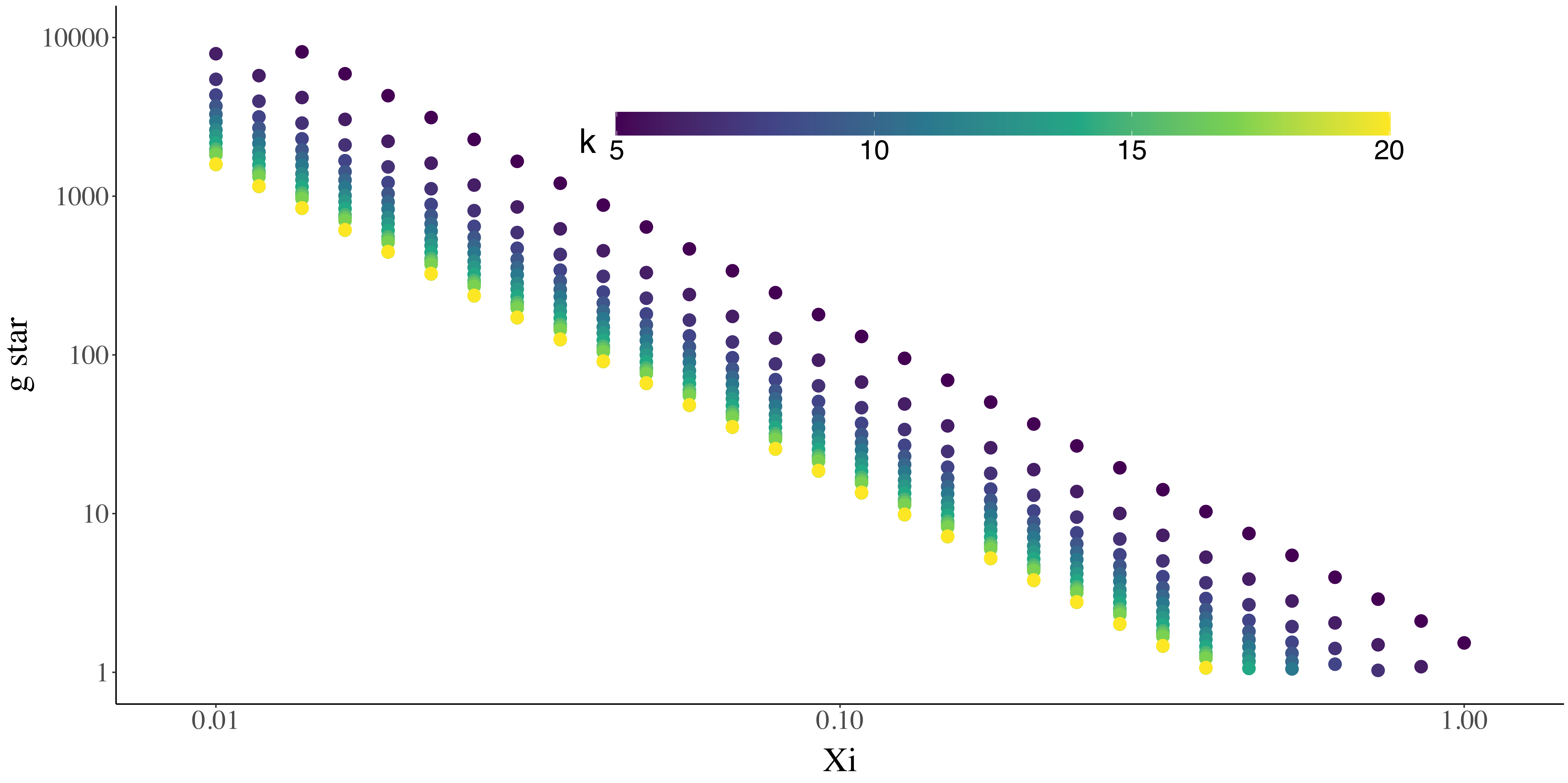
Outline

1. Proposal and Generic Validity
2. Non-Asymptotic Theory
   (i) Concentration and Normal Approximation
   (ii) Reproducibility
3. Performance

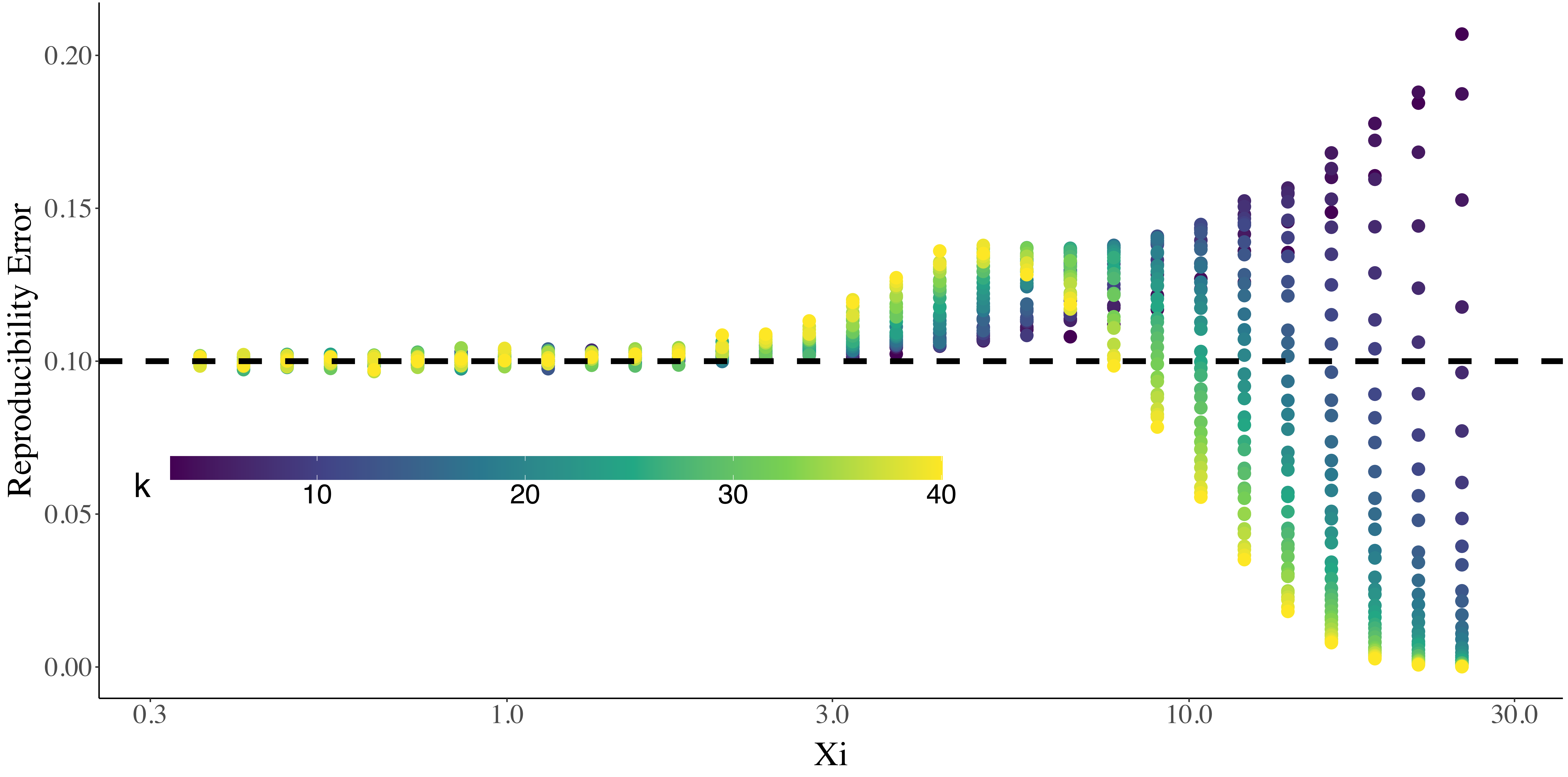
Treatment Effect Estimation : Reproducibility Error $P\{\,|a(\mathsf{R}_{\hat{g},k},D)-a(\mathsf{R}'_{\hat{g}',k},D)|\leq\xi\mid D\}$



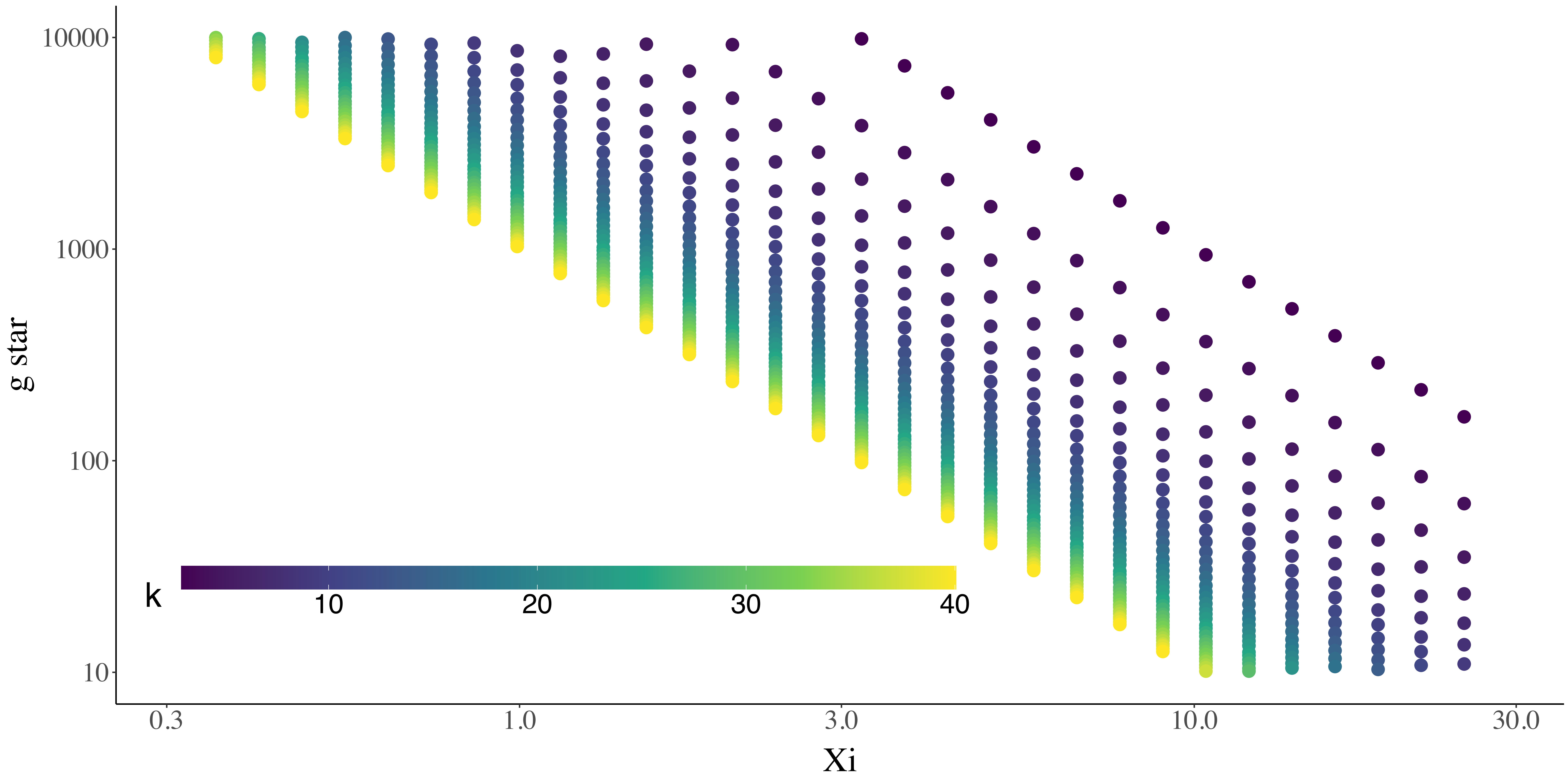
Treatment Effect Estimation: Oracle Stopping time $g^\star = 2v_{1,k}(D)(z_{1-\beta/2}/\xi)$



Risk Estimation: Reproducibility Error $P\{\,|a(\mathsf{R}_{\hat{g},k},D)-a(\mathsf{R}'_{\hat{g}',k},D)|\leq \xi \mid D\}$



Risk Estimation: Oracle Stopping time $g^\star = 2v_{1,k}(D)(z_{1-\beta/2}/\xi)$



Conclusion

- We propose a method for sequentially aggregating sample-split statistics to ensure that residual randomness is small
- We give two main results:
  - Cross-splitting reduces randomness more quickly than independent splitting
  - But does not necessary improve the quality of the nominal error rate
- Consequence: Users navigate tradeoff between computation and accuracy