Median bias, HulC, and Valid Inference

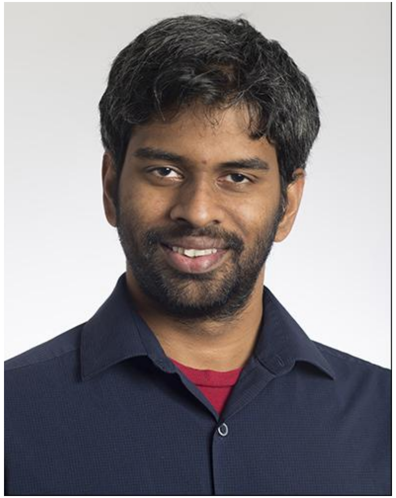
Arun Kumar Kuchibhotla

Carnegie Mellon University

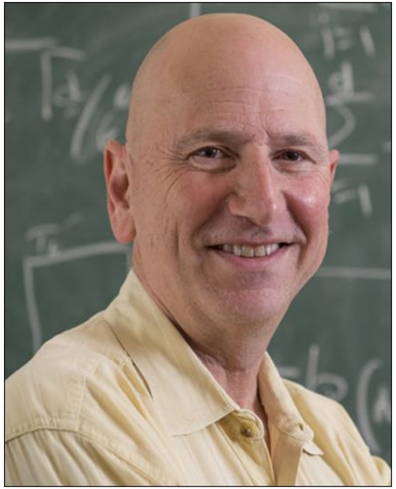
https://arxiv.org/abs/2105.14577
https://arxiv.org/abs/2106.00164

Collaborators

Sivaraman Balakrishnan



Larry Wasserman



Introduction

- Confidence interval is one of the key components of statistical inference.
- Traditional methods of inference are based on the (limiting) distribution of a point estimator.
- There are two "general" methods for construction of confidence intervals:
  - Wald technique: Estimating parameters (e.g., variance) of limiting distribution and using the quantiles of the limiting distribution;
  - Resampling techniques: Estimate the limiting distribution by resampling data and then use quantiles of the estimated distribution.
- Limiting distribution is also crucially used in defining regularity of an estimator and this in turn is used for discussing uniformly valid inference.

Outline

- ❖ Median bias
- ❖ HulC
- ❖ Simulation Examples
- ❖ Valid Inference

Median bias: Introduction

Median bias of an estimator

An estimator $\hat{\theta}_n$ as a function of the data is said to be median unbiased for θ_0 if $\text{Median}(\hat{\theta}_n) = \theta_0$, that is

$$\min\left\{\mathbb{P}(\hat{\theta}_n \leq \theta_0),\, \mathbb{P}(\hat{\theta}_n \geq \theta_0)\right\} \;\geq\; \tfrac{1}{2}.$$

In words, this means that the estimator both over- and under-estimates θ_0 with a probability of at least ½.

If the estimator is equal to the target almost surely, then both the probabilities are one and the estimator is considered median unbiased.

Median bias of an estimator

An estimator $\hat{\theta}_n$ as a function of the data is said to be median unbiased for θ_0 if $\text{Median}(\hat{\theta}_n) = \theta_0$, that is

$$\min\left\{\mathbb{P}(\hat{\theta}_n \le \theta_0),\, \mathbb{P}(\hat{\theta}_n \ge \theta_0)\right\} \;\ge\; \tfrac{1}{2}.$$

In words, this means that the estimator both over- and under-estimates θ_0 with a probability of at least ½.

In general, we define the median bias of an estimator with respect to a functional as

$$\text{Med-bias}_{\theta_0}(\hat{\theta}_n) \;:=\; \left(\tfrac{1}{2} - \min\left\{\mathbb{P}(\hat{\theta}_n \le \theta_0),\, \mathbb{P}(\hat{\theta}_n \ge \theta_0)\right\}\right)_+.$$

Median bias: Examples

$$\text{Med-bias}_{\theta_0}(\hat{\theta}_n) \;:=\; \left(\tfrac{1}{2} - \min\left\{\mathbb{P}(\hat{\theta}_n \le \theta_0),\, \mathbb{P}(\hat{\theta}_n \ge \theta_0)\right\}\right)_+.$$

1. Suppose $X_1, \ldots, X_n \overset{iid}{\sim} N(\theta_0, 1)$, then $\hat{\theta}_n = \overline{X}_n$ is median unbiased. The same holds for any symmetric location family.
2. Suppose $X_1, \ldots, X_n$ are iid with median θ_0, then
$$\hat{\theta}_n = \begin{cases} X_{(r)}, & \text{with probability } 1/2, \\ X_{(n-r+1)}, & \text{with probability } 1/2, \end{cases}$$
is median unbiased.
3. Suppose $X_1, \ldots, X_n \overset{iid}{\sim} \text{Unif}(0, \theta_0)$, then $\hat{\theta}_n = 2X_{(n)} - X_{(n-1)}$ is median unbiased. The largest order statistic has the largest median bias of ½.
4. Suppose $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, 1)$, then $\hat{\theta}_n = \overline{X}_n \mathbf{1}\{\overline{X}_n \ge 0\}$ is median unbiased for $\theta_0 = \mu \mathbf{1}\{\mu \ge 0\}$. Same for symmetric location family.

Median bias: Examples (Contd.)

An estimator $\hat{\theta}_n$ is said to be *asymptotically* median unbiased if

$$\lim_{n \to \infty} \text{Med-bias}_{\theta_0}(\hat{\theta}_n) = 0.$$

1. Any estimator with a limiting normal distribution after proper normalization is asymptotically median unbiased.

   This includes asymptotically linear estimators considered in the efficiency framework of parametric and semi-/non-parametric models.

2. Any estimator with a limiting distribution symmetric around zero after proper normalization is asymptotically median unbiased.

   This includes examples from shape constrained literature where the limiting distribution is non-standard, e.g., Chernoff distribution.

Some comments on Median bias

An estimator $\hat{\theta}_n$ is said to be *asymptotically* median unbiased if

$$\lim_{n\to\infty} \text{Med-bias}_{\theta_0}(\hat{\theta}_n) = 0.$$

1. No limiting distribution of the estimator is required to establish its median bias properties.

   E.g.: If $\hat{\theta}_n$ solves $\mathbb{Z}_n(\theta) = 0$ for some differentiable moment equation, then

$$\hat{\theta}_n - \theta_0 = \frac{\mathbb{Z}_n(\theta_0)}{\mathbb{Z}_n'(\tilde{\theta}_n)}.$$

$$\min\{\mathbb{P}(\hat{\theta}_n \le \theta_0), \mathbb{P}(\hat{\theta}_n \ge \theta_0) = \min\{\mathbb{P}(\mathbb{Z}_n(\theta_0) \ge 0), \mathbb{P}(\mathbb{Z}_n(\theta_0) \le 0).$$

$$\text{Median unbiased if } \mathbb{Z}_n(\theta_0) \xrightarrow{d} L \text{ with } L \stackrel{d}{=} -L.$$

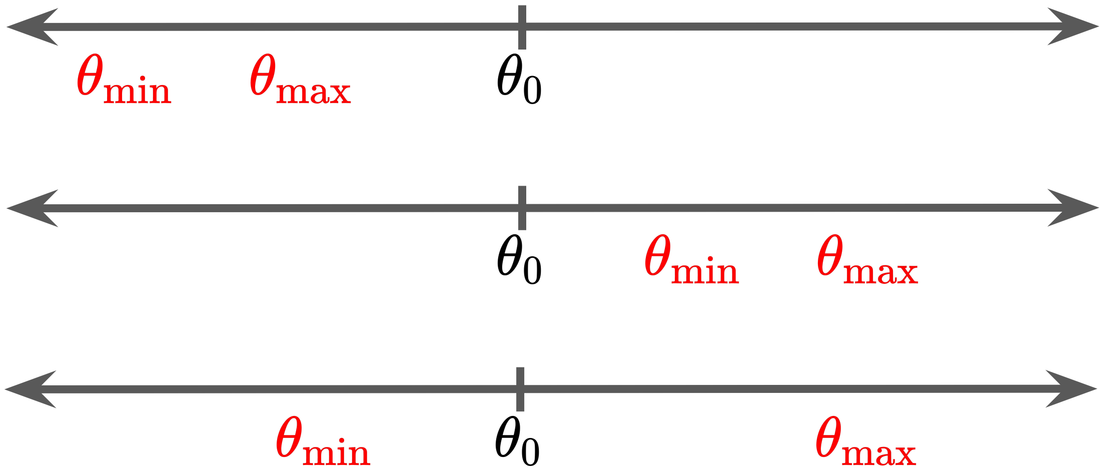
Introducing The HulC

Hull based **C**onfidence

Motivating Calculations

If we have two indep. estimators $\hat{\theta}^{(1)}, \hat{\theta}^{(2)}$, median unbiased for θ_0

$$\theta_{\min} = \min\{\hat{\theta}^{(1)}, \hat{\theta}^{(2)}\}, \quad \text{and} \quad \theta_{\max} = \max\{\hat{\theta}^{(1)}, \hat{\theta}^{(2)}\}.$$



$$\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

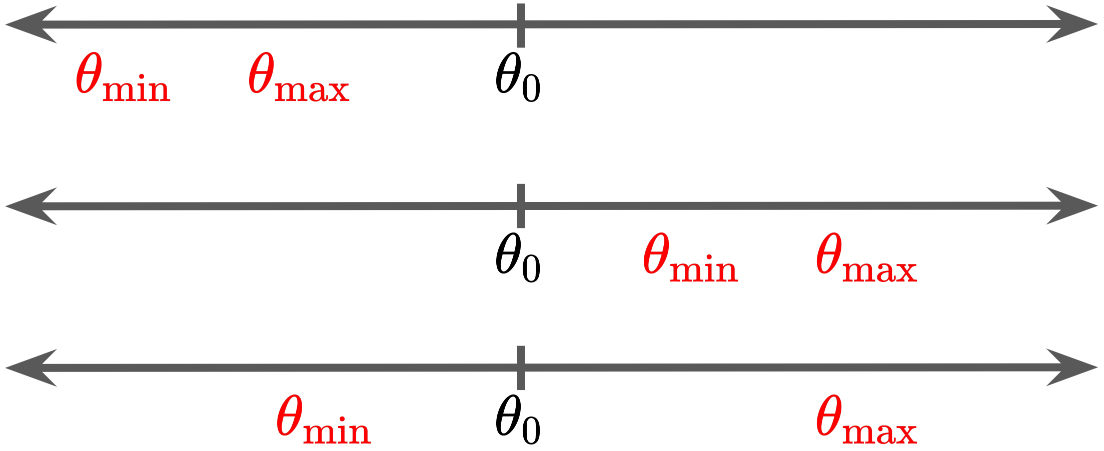
$$\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

$$\frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2}$$

Motivating Calculations

If we have B indep. estimators $\hat{\theta}^{(j)}$, $1 \leq j \leq B$, median unbiased about θ_0

$$\theta_{\min} = \min\{\hat{\theta}^{(j)} : j \leq B\}, \quad \text{and} \quad \theta_{\max} = \max\{\hat{\theta}^{(j)} : j \leq B\}.$$



2^{-B}

2^{-B}

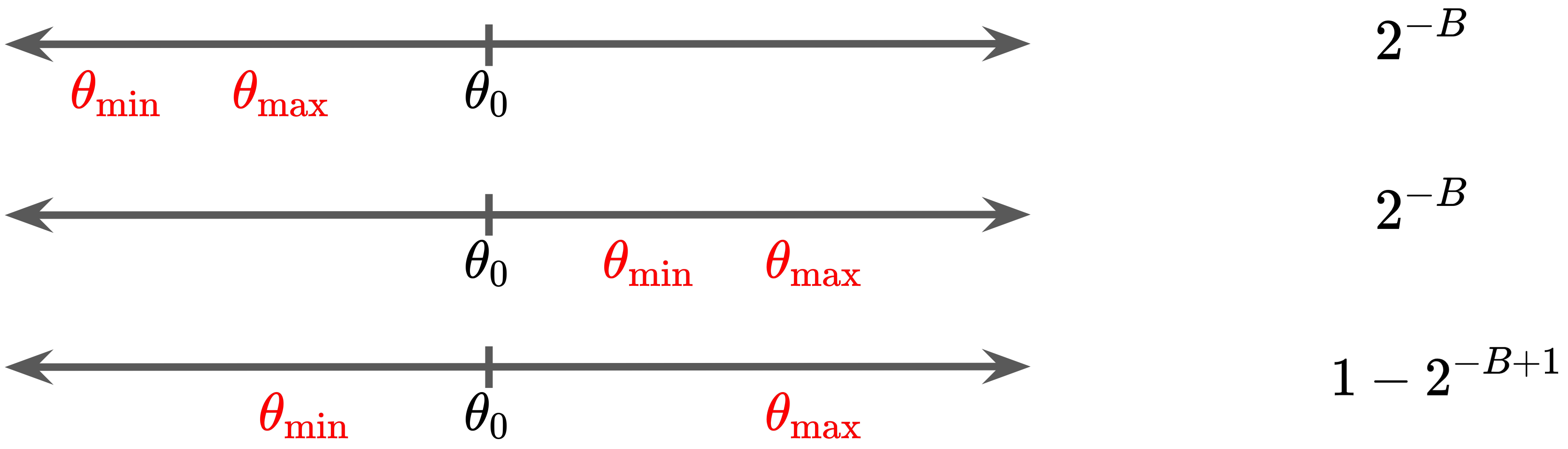
$1 - 2^{-B+1}$

No Reference to the rate of convergence

Motivating Calculations

If we have B indep. estimators $\hat{\theta}^{(j)}, 1 \leq j \leq B$, median unbiased about θ_0

$$\theta_{\min} = \min\{\hat{\theta}^{(j)} : j \leq B\}, \quad \text{and} \quad \theta_{\max} = \max\{\hat{\theta}^{(j)} : j \leq B\}.$$



$$\text{If } B = B_\alpha = \lceil \log_2(2/\alpha) \rceil, \text{ then } 1 - 2^{-B+1} \geq 1 - \alpha.$$

$B_\alpha = \lceil \log_2(2/\alpha) \rceil$

General result

If we have independent estimators $\hat{\theta}_j, 1 \le j \le B_\alpha$ and

$$\Delta_{n,\alpha} := \max_{1 \le j \le B_\alpha} \left(\tfrac{1}{2} - \min\left\{ \mathbb{P}(\hat{\theta}_j \le \theta_0), \mathbb{P}(\hat{\theta}_j \ge \theta_0) \right\} \right)_+,$$

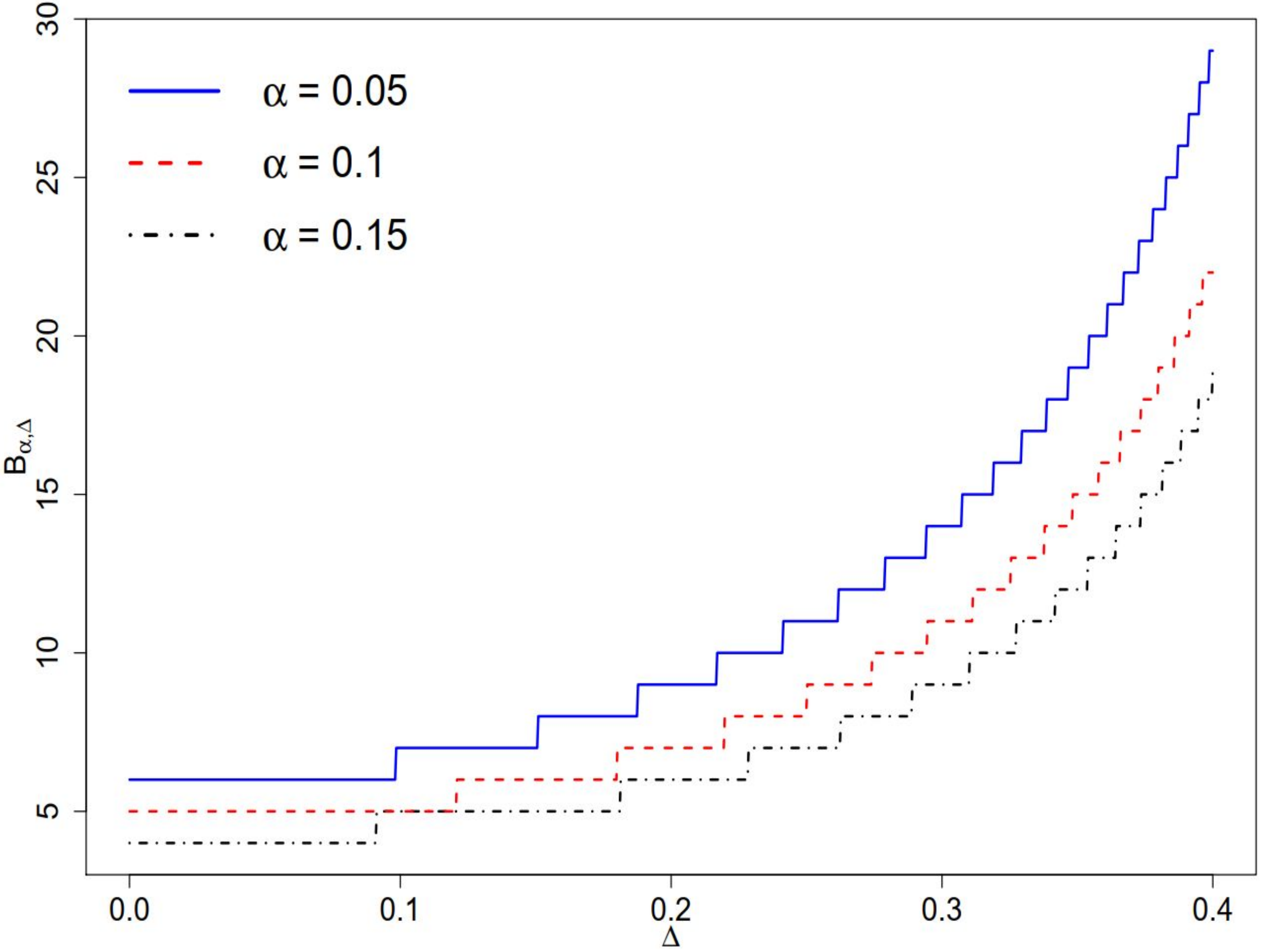
then

$$\begin{aligned} \mathbb{P}\left(\theta_0 \notin [\theta_{\min}, \theta_{\max}]\right) &\le P(B_\alpha; \Delta_{n,\alpha}) \\ &= \left(\frac{1}{2} - \Delta_{n,\alpha}\right)^{B_\alpha} + \left(\frac{1}{2} + \Delta_{n,\alpha}\right)^{B_\alpha}. \end{aligned}$$

Independent copies of the estimator can be obtained by splitting the data.
So, each estimator can be based on $n / \lceil \log_2(2/\alpha) \rceil$ observations.

If $\Delta_{n,\alpha} = o(1)$, then the miscoverage is asymptotically less than α.

The number of splits *increase* as the median bias of the estimators *increase*.



Coverage/width of HulC

$B_\alpha = \lceil \log_2(2/\alpha) \rceil$

Coverage of HulC

If we have independent estimators $\hat{\theta}_j, 1 \le j \le B_\alpha$ and

$$\Delta_{n,\alpha} := \max_{1\le j\le B_\alpha} \left(\frac{1}{2} - \min\left\{\mathbb{P}(\hat{\theta}_j \le \theta_0),\, \mathbb{P}(\hat{\theta}_j \ge \theta_0)\right\}\right)_+,$$

then

$$\begin{aligned}\mathbb{P}\left(\theta_0 \notin [\theta_{\min},\, \theta_{\max}]\right) &\le P(B_\alpha; \Delta_{n,\alpha}) \\ &= \left(\frac{1}{2} - \Delta_{n,\alpha}\right)^{B_\alpha} + \left(\frac{1}{2} + \Delta_{n,\alpha}\right)^{B_\alpha}.\end{aligned}$$

A Taylor series expansion yields

0

$$P(B_\alpha; \Delta_{n,\alpha}) = P(B_\alpha; 0) + P'(B_\alpha; 0)\Delta_{n,\alpha} + O(B_\alpha^2 \Delta_{n,\alpha}^2).$$

$B_\alpha = \lceil \log_2(2/\alpha) \rceil$

Coverage of HulC (cont.)

If the estimators $\hat{\theta}_j, 1 \le j \le B_\alpha$ are independent, then

$$\mathbb{P}\left(\theta_0 \notin \left[\min_{1\le j\le B_\alpha} \hat{\theta}_j,\ \max_{1\le j\le B_\alpha} \hat{\theta}_j\right]\right) \le \alpha\left(1 + CB_\alpha^2 \Delta_{n,\alpha}^2\right).$$

Hence, the existence of an asymptotically median unbiased estimator implies the existence of an asymptotic

Multiplicative error
Second-order accuracy

If $\Delta_{n,\alpha} = O(\sqrt{\log_2(2/\alpha)/n})$, then

$$\mathbb{P}\left(\theta_0 \notin \left[\min_{1\le j\le B_\alpha} \hat{\theta}_j,\ \max_{1\le j\le B_\alpha} \hat{\theta}_j\right]\right) \le \alpha\left(1 + C\frac{(\log_2(2/\alpha))^3}{n}\right).$$

Wald and bootstrap confidence intervals have a first-order coverage.
Second-order correct bootstrap intervals exist.

Some notes on coverage

- For asymptotically symmetric estimators, the HulC confidence intervals are second-order accurate.
- This second-order accuracy holds for estimators that do not even converge in distribution (e.g., Z-estimators).
- Depending only on median bias, HulC operates under weaker conditions than bootstrap and subsampling (WHY?).
- With estimators with reduced median bias, the HulC confidence intervals are sixth-order accurate (i.e., coverage error of n^{-3}).

Some notes on coverage (Contd.)

If $\Delta_{n,\alpha} = O(\sqrt{\log_2(2/\alpha)/n})$, then

$$\mathbb{P}\left(\theta_0 \notin \left[\min_{1\le j\le B_\alpha} \hat{\theta}_j,\ \max_{1\le j\le B_\alpha} \hat{\theta}_j\right]\right) \le \alpha\left(1 + C\frac{(\log_2(2/\alpha))^3}{n}\right).$$

Even as α tends to zero, the HulC interval attains (relative) miscoverage unlike Wald and bootstrap intervals.

α is allowed to converge to zero almost exponentially in the sample size. This helps in controlling coverage for problems in high-dimensional statistics or multiple testing via union bound.

Width of HulC Intervals

If the estimators are asymptotically normal with $n^{1/2}$ rate, then the $(1-\alpha)$ confidence HulC and Wald intervals satisfy (asymptotically)

$$\frac{\text{Width of HulC}}{\text{Width of Wald}} = \sqrt{\log_2(\log_2(2/\alpha))} \geq 1.$$

The ratio of widths is larger than one but grows very slowly as $\alpha \to 0.$

In a way, this is the price to pay for the generality of HulC.

HulC does not estimate variance or unknown parameters of the limiting distribution.

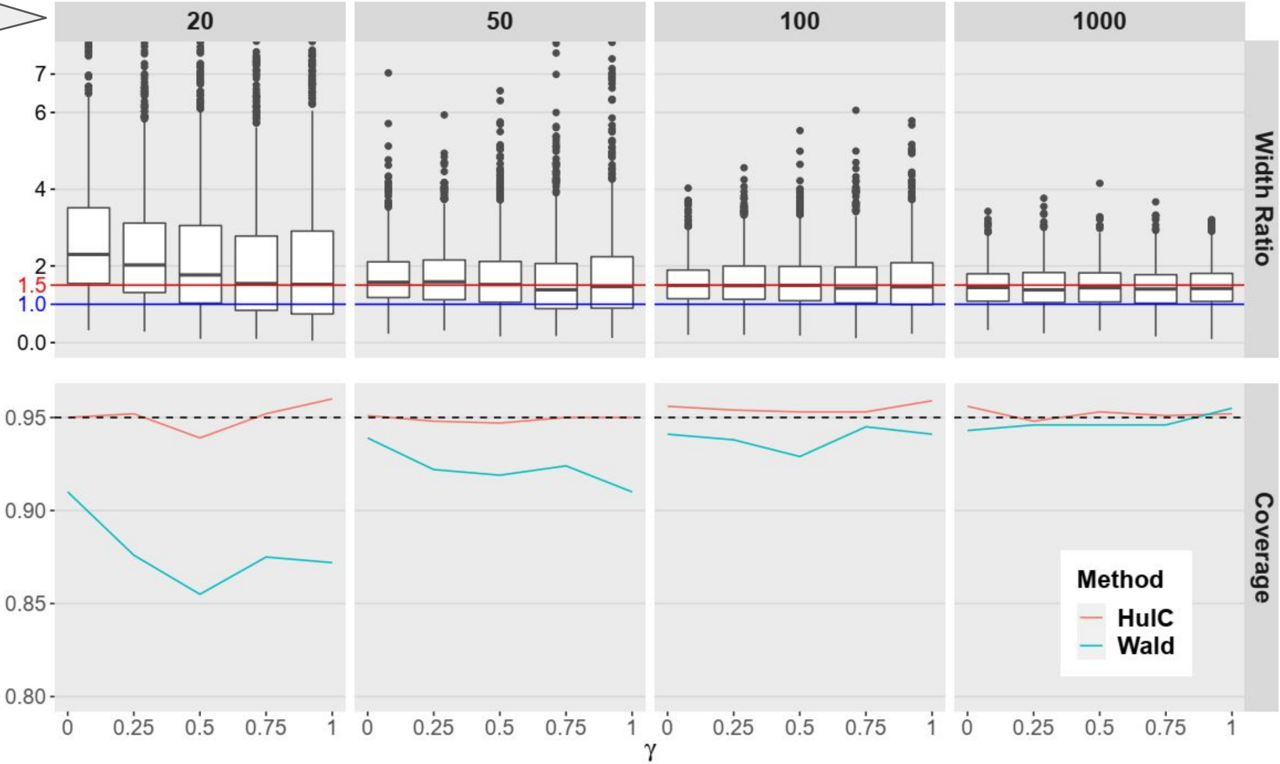
Simulation Examples

OLS Linear Regression

$$Y_i = 1 + 2X_i + \gamma X_i^{1.7} + \exp(\gamma X_i)\xi_i$$

$$X_i \sim \text{Unif}[0, 10], \quad \xi_i \sim N(0, 1).$$

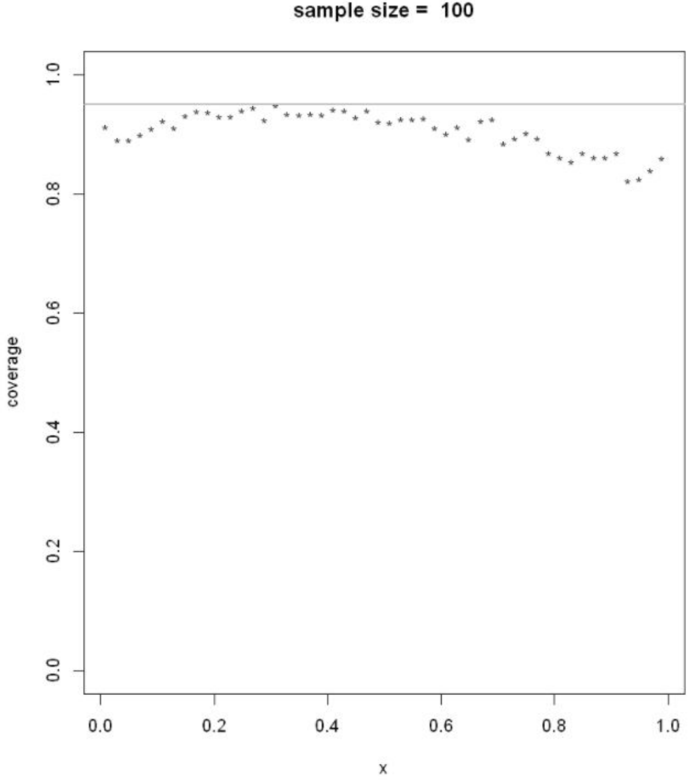
Sample size

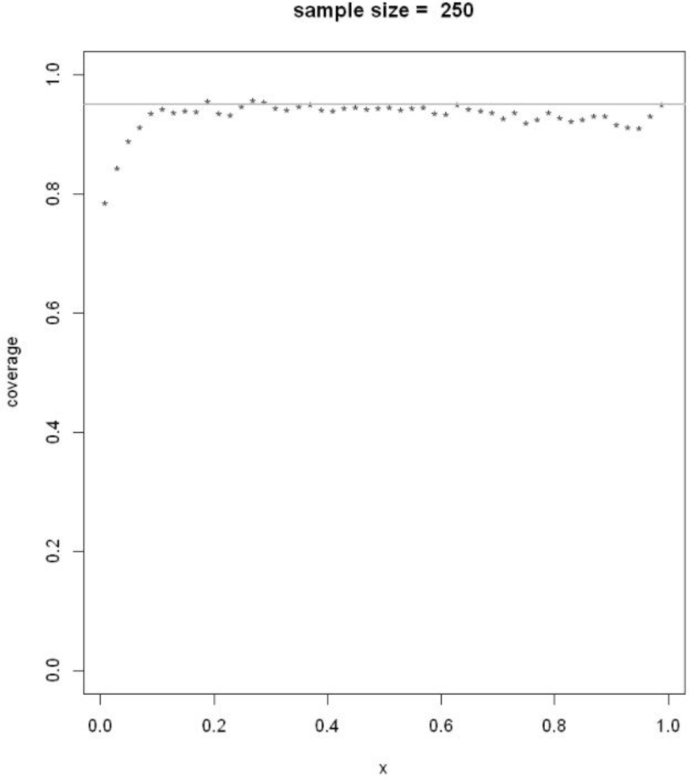


Monotone Regression: LSE

$$Y_i = \exp(2X_i) + \xi_i,$$
$$X_i \sim \text{Unif}[0,1],\ \xi_i | X_i \sim N(0,1)$$

$$\widehat{f} := \underset{f:\text{non-decreasing}}{\text{argmin}} \sum_{i=1}^{n} (Y_i - f(X_i))^2.$$





Valid Inference

Assumptions of HulC

- HulC has very nice properties when the underlying estimator of the functional is asymptotically median unbiased.
- HulC gives multiplicative miscoverage and second order accuracy with ease and works even in high-dimensional problems.
- But HulC seems *weaker* than bootstrap, subsampling, and other inferential methods because it restricts itself to asymptotically median unbiased estimators.
- Not all functionals may have asymp. median unbiased estimators, but

Assumptions of HulC

- HulC has very nice properties when the underlying estimator of the functional is asymptotically median unbiased.
- HulC gives multiplicative miscoverage and second order accuracy with ease and works even in high-dimensional problems.
- But HulC seems *weaker* than bootstrap, subsampling, and other inferential methods because it restricts itself to asymptotically median unbiased estimators.
- Not all functionals may have asymp. median unbiased estimators, but

$$\text{Inference} \Leftrightarrow \text{Asymp. Median Unbiased} \Leftrightarrow \text{HulC}$$

Inference and Median Unbiasedness

For a functional/parameter θ_0 has an asymptotically valid confidence interval at some level based on IID data

if and only if

there exists an asymptotically median unbiased estimator for it.

Inference and Median Unbiasedness

Fix any set of distributions $\mathcal{P}_n$, any functional $\theta : \mathcal{P}_n \to \mathbb{R}$ and any $\gamma \in (0,1)$.

There exists a confidence interval procedure satisfying

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}_n} \mathbb{P}_P(\theta(P) \notin \widehat{\mathrm{CI}}_{\gamma,n}) \le \gamma,$$

if and only if

there exists an estimator sequence $\hat{\theta}_n,\ n \ge 1$ satisfying

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}_n} \left(\frac{1}{2} - \min\left\{\mathbb{P}_P(\hat{\theta}_n \le \theta(P)),\, \mathbb{P}_P(\hat{\theta}_n \ge \theta(P))\right\}\right)_+ = 0.$$

Inference and Median Unbiasedness

Fix any set of distributions $\mathcal{P}_n$, any functional $\theta : \mathcal{P}_n \to \mathbb{R}$ and any $\gamma \in (0, 1)$.

There exists a confidence interval procedure satisfying

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}_n} \mathbb{P}_P(\theta(P) \notin \widehat{\mathrm{CI}}_{\gamma,n}) \le \gamma,$$

if and only if

there exists an estimator sequence $\hat{\theta}_n$, $n \ge 1$ satisfying

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}_n} \left(\frac{1}{2} - \min\left\{\mathbb{P}_P(\hat{\theta}_n \le \theta(P)), \mathbb{P}_P(\hat{\theta}_n \ge \theta(P))\right\}\right)_+ = 0.$$

Median Regular Estimator

Conclusions

- HulC is a general purpose inference method like bootstrap and subsampling.
- Unlike classical methods, HulC does not depend on convergence in distribution but on median bias.
- For asymptotically normal estimators, HulC yields second-order accurate confidence intervals.
- Median bias required for HulC can be estimated via subsampling leading to adaptive HulC.
- Median bias control is necessary for valid inference. It is a necessary regularity notion for honest inference.

Thanks for your attention!!

Future Research

- Is there a systematic way to construct asymptotically median unbiased estimators?
- For asymptotically median unbiased estimators, are there better inference methods than HulC?
    - Yes, but what is the complete class? What’s the best?