The HulC

Hull based Confidence Regions

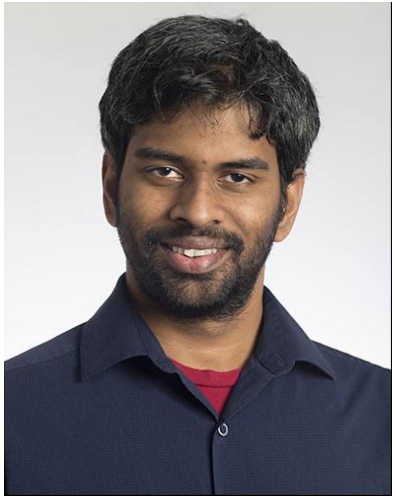
https://arxiv.org/abs/2105.14577

Arun Kumar Kuchibhotla

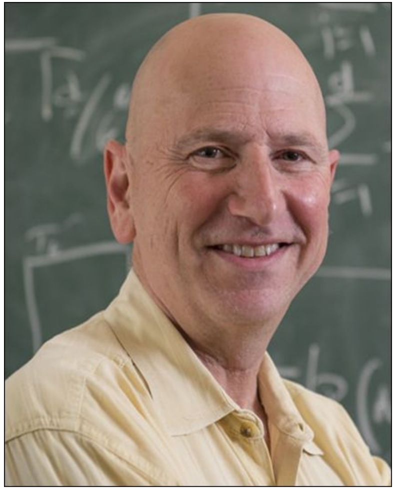
Carnegie Mellon University

Collaborators

Sivaraman Balakrishnan



Larry Wasserman



Introduction

- Confidence interval is one of the key components of statistical inference.
- There are three "general" methods for construction of confidence intervals:
  - Estimating parameters (e.g., variance) of limiting distribution;
  - Bootstrap;
  - Subsampling.

Introducing a new general-purpose method: HulC

Outline

- ❖ HulC
- ❖ Coverage and width properties
- ❖ Simulation Examples
- ❖ Adaptive HulC

Introducing The HulC

Motivation

- Suppose $\hat{\theta}$ is a consistent estimator of $\theta_0 \in \mathbb{R}$.
- With $\hat{\theta}^{(1)}, \hat{\theta}^{(2)}, \ldots,$ representing independent copies of $\hat{\theta}$, we *expect* that the minimum and maximum of the estimators contain θ_0.
- This will hold if the support of the distribution of $\hat{\theta}$ contains as an interior point θ_0.

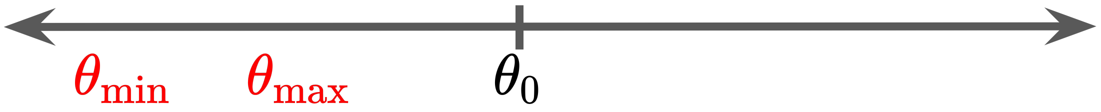
Problem: When does this hold true? And how many estimators are needed for valid coverage?

Answer: The estimator is not pathologically “asymmetric.”

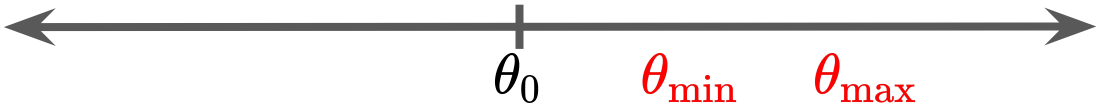
Simple Calculations

Suppose we have two estimators $\hat{\theta}^{(1)}, \hat{\theta}^{(2)}$, symmetric about θ_0

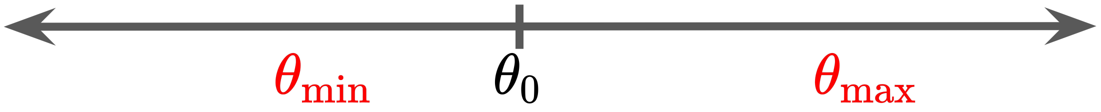
$$\theta_{\min} = \min\{\hat{\theta}^{(1)}, \hat{\theta}^{(2)}\}, \quad \text{and} \quad \theta_{\max} = \max\{\hat{\theta}^{(1)}, \hat{\theta}^{(2)}\}.$$



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



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

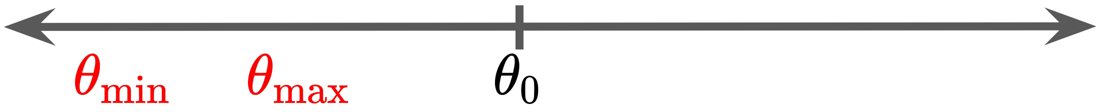


$$\frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2}$$

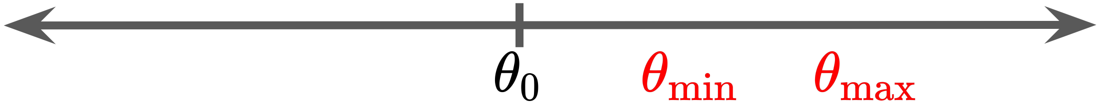
Simple Calculations

Suppose we have B estimators $\hat{\theta}^{(j)}, 1 \le j \le B,$ symmetric about θ_0

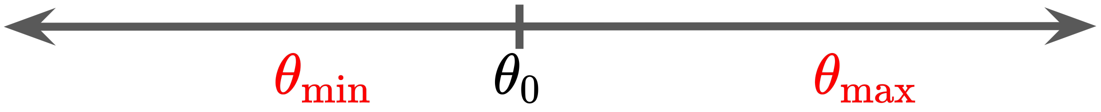
$$\theta_{\min} = \min\{\hat{\theta}^{(j)} : j \le B\}, \quad \text{and} \quad \theta_{\max} = \max\{\hat{\theta}^{(j)} : j \le B\}.$$



$$\frac{1}{2^B}$$



$$\frac{1}{2^B}$$



$$1 - \frac{2}{2^B}$$

Some notes

- We don't need the estimators $\hat{\theta}^{(j)}, 1 \le j \le B,$ to be symmetric around θ_0
- It suffices to have

$$\mathbb{P}(\hat{\theta}^{(j)} \le \theta_0) \ge \tfrac{1}{2} \quad \text{and} \quad \mathbb{P}(\hat{\theta}^{(j)} \ge \theta_0) \ge \tfrac{1}{2}.$$

- This is equivalent to

$$\text{Median}(\hat{\theta}^{(j)}) = \theta_0.$$

- With this condition again, we get that θ_0 lies in the minimum to the maximum of B estimators with a probability of at least $1 - 2^{-B+1}$.
- The calculation can be extended using median bias of the estimator

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

General result

If

Median bias of
the estimators.

$$\Delta := \max_{j \geq 1} \left(\frac{1}{2} - \min \left\{ \mathbb{P}(\hat{\theta}_j \leq \theta_0), \mathbb{P}(\hat{\theta}_j \geq \theta_0) \right\} \right)_+ ,$$

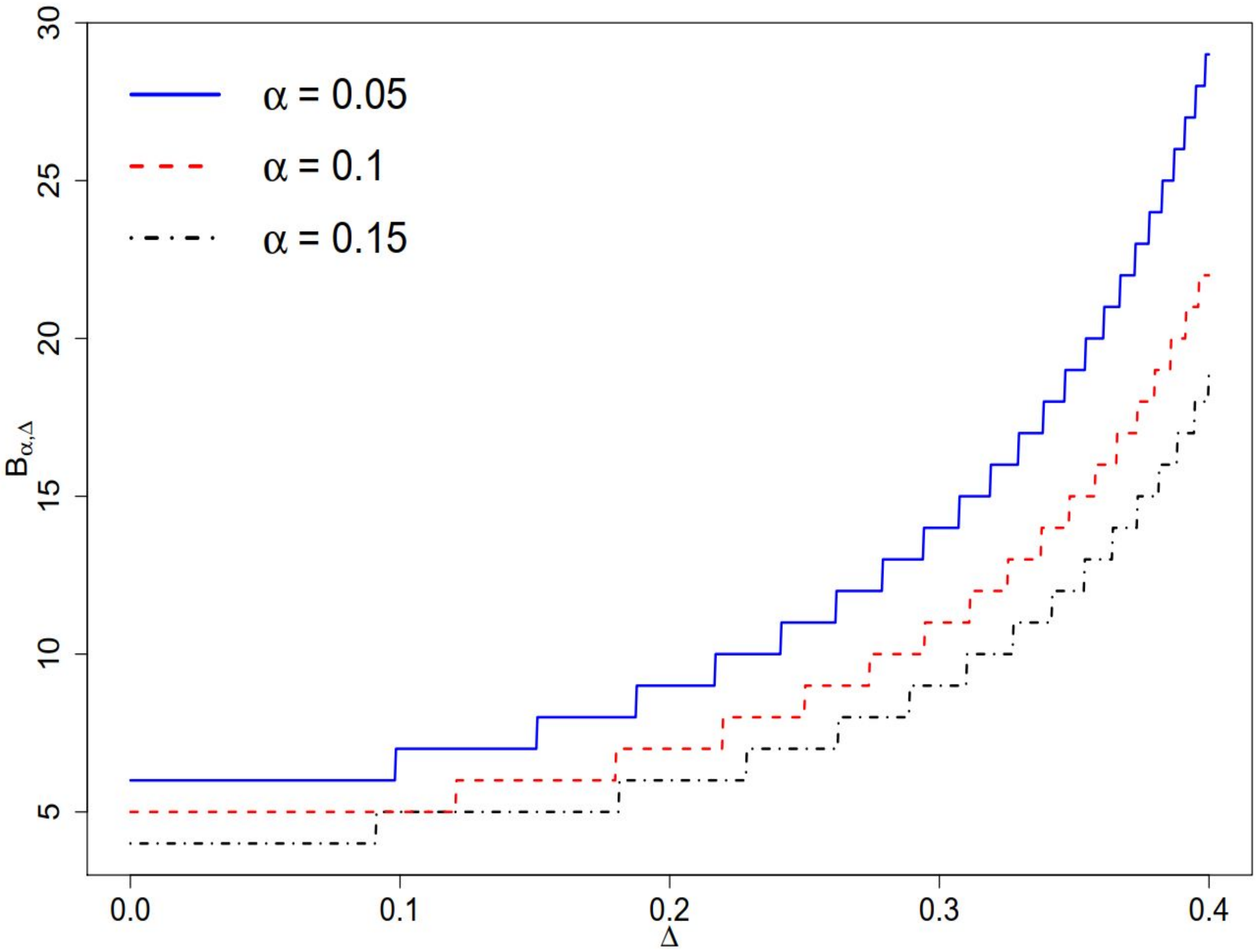
then

$$\mathbb{P}\left(\theta_0 \notin [\theta_{\min}, \theta_{\max}]\right) \leq P(B; \Delta)$$
$$= \left(\frac{1}{2} - \Delta\right)^B + \left(\frac{1}{2} + \Delta\right)^B .$$

Hence, $B = B_{\alpha,\Delta}$ with $P(B;\Delta) \leq \alpha$ many estimators to get a coverage of at least $1 - \alpha$.

Independent copies of the estimator can be obtained by splitting the data.

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



The quantity Δ is the maximum of the median biases of $\hat{\theta}^{(j)}, 1 \leq j \leq B,$

$$\Delta = \max_{j \geq 1} \text{Med-bias}_{\theta_0}(\hat{\theta}^{(j)}).$$

If we take $\theta_0 := \text{Median}(\hat{\theta}^{(j)})$, then with $B = \lceil \log_2(2/\alpha) \rceil,$

$$\mathbb{P}\left(\theta_0 \notin \left[\min_{1 \leq j \leq B} \hat{\theta}^{(j)}, \max_{1 \leq j \leq B} \hat{\theta}^{(j)}\right]\right) \leq \alpha.$$

Unfortunately, the median of the estimators is not our target, but the limit of the estimators is.

In some cases, it is possible to directly bound the difference between the median and the limit of the estimators.

Application to Binomial Proportion

$$X_1, X_2, \ldots, X_n \sim \text{Bernoulli}(p)$$

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

$$\hat{\theta}^{(j)} = \frac{1}{|S_j|} \sum_{i \in S_j} X_i, \text{ the sample mean on } j\text{-th batch.}$$

Note that $\hat{\theta}^{(j)}$ are identically distributed as $\text{Binomial}(n/B_\alpha, p)/(n/B_\alpha)$.

$$\left| \text{Median}\left(\frac{\text{Binom}(n/B_\alpha, p)}{n/B_\alpha} \right) - p \right| \leq \frac{B_\alpha \log(2)}{n} \quad \text{for all} \quad p \in [0, 1].$$

$$\Rightarrow \quad \mathbb{P}\left(p \notin \left[\min_{1 \leq j \leq B_\alpha} \hat{\theta}_j - \frac{B_\alpha \log 2}{n}, \max_{1 \leq j \leq B_\alpha} \hat{\theta}_j + \frac{B_\alpha \log 2}{n} \right] \right) \leq \alpha.$$

Coverage and Width

Coverage of HulC

If

$$\Delta := \max_{j \geq 1} \left(\frac{1}{2} - \min \left\{ \mathbb{P}(\hat{\theta}_j \leq \theta_0), \mathbb{P}(\hat{\theta}_j \geq \theta_0) \right\} \right)_+,$$

then

$$\begin{aligned} \mathbb{P}\left(\theta_0 \notin \left[\min_{1 \leq j \leq B} \hat{\theta}_j, \max_{1 \leq j \leq B} \hat{\theta}_j \right] \right) &\leq P(B; \Delta) \\ &= \left(\frac{1}{2} - \Delta \right)^B + \left(\frac{1}{2} + \Delta \right)^B. \end{aligned}$$

Take $B = B_{\alpha,0}$ with $P(B;0) \leq \alpha$. A Taylor series expansion yields

0

$$P(B; \Delta) = P(B; 0) + P'(B; 0)\Delta + O(\Delta^2).$$

Coverage of HulC (cont.)

If

$$\Delta := \max_{j \geq 1} \left(\frac{1}{2} - \min \left\{ \mathbb{P}(\hat{\theta}_j \leq \theta_0), \, \mathbb{P}(\hat{\theta}_j \geq \theta_0) \right\} \right)_+ ,$$

and $\Delta = O(n^{-1/2})$, then

$$\mathbb{P}\left(\theta_0 \notin \left[\min_{1 \leq j \leq B} \hat{\theta}_j, \, \max_{1 \leq j \leq B} \hat{\theta}_j \right] \right) \leq \alpha \left(1 + O(n^{-1}) \right).$$

Multiplicative error
Second-order accuracy (n^{-1})

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.
- Depending only on median bias, HulC operates under weaker conditions than bootstrap and subsampling.
- With estimators with reduced median bias, the HulC confidence intervals are sixth-order accurate (i.e., coverage error of n^{-3}).

Inference if and only if HulC

Estimators with a specific limiting median bias is important for applying HulC. Existence of limiting distribution implies this.

Interestingly, if there exists asymptotically valid confidence intervals, then there exists asymptotically median unbiased estimators. If there is a procedure leading to confidence intervals such that

$$\sup_{\gamma \in [0,1]} \left| \mathbb{P}(\theta_0 \notin \widehat{\mathbf{CI}}_\gamma) - \gamma \right| \leq r_n \to 0,$$

then there exists an estimator $\hat{\theta}$ such that

$$\text{Med-bias}_{\theta_0}(\hat{\theta}) \leq r_n.$$

Hence, HulC applies with this estimator and yields a confidence interval with multiplicative miscoverage error and second order accuracy.

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.

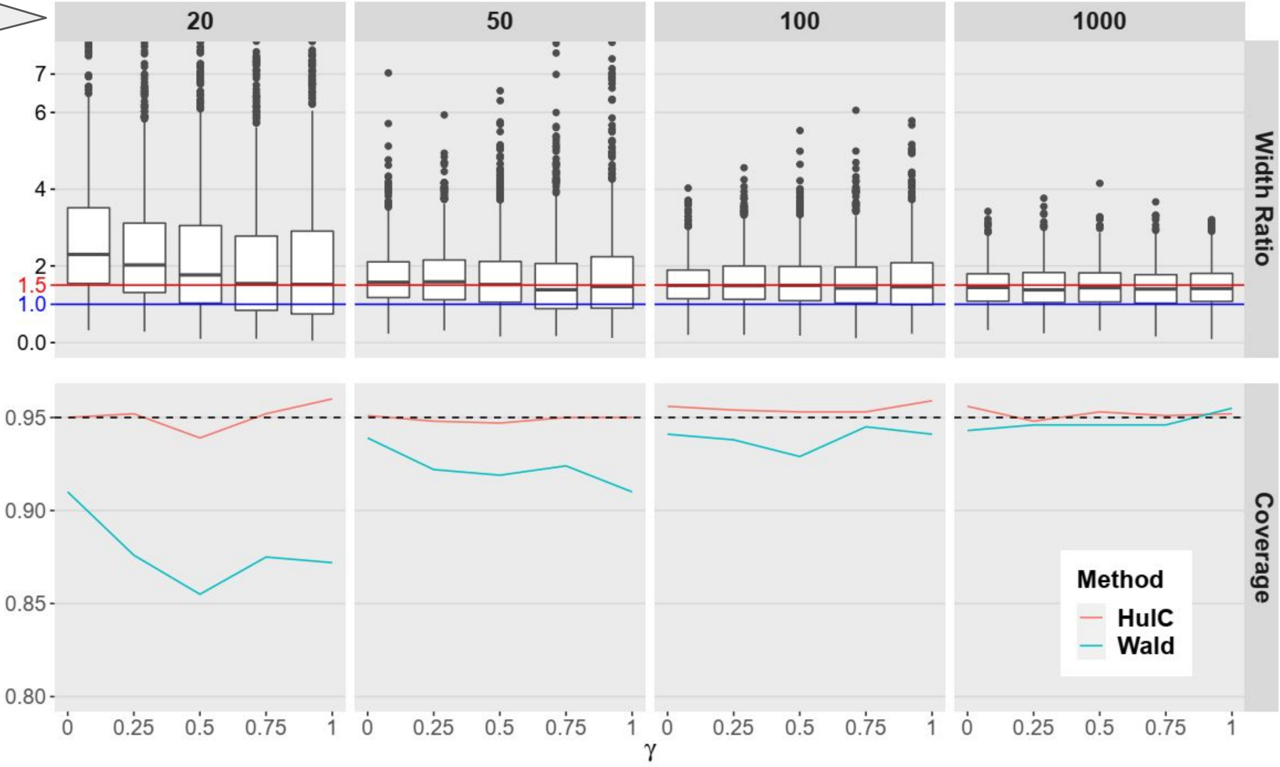
Simulation Examples

OLS Linear Regression

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

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

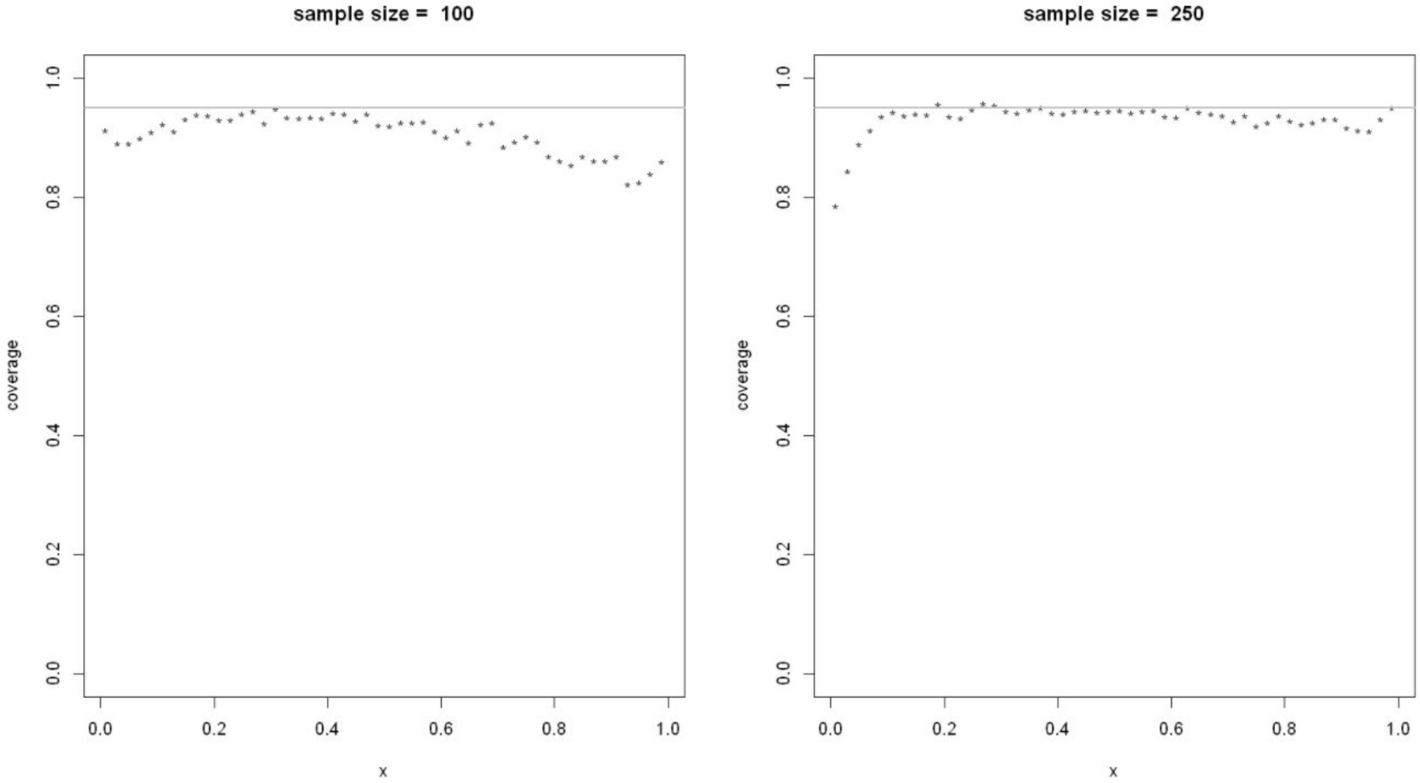
Sample size



Monotone Regression: LSE

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

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



Adaptive HulC

Recall: coverage of HulC

If

$$\Delta := \max_{j \geq 1} \left(\tfrac{1}{2} - \min\left\{ \mathbb{P}(\hat{\theta}_j \leq \theta_0),\ \mathbb{P}(\hat{\theta}_j \geq \theta_0) \right\} \right)_+,$$

then

$$\begin{aligned} \mathbb{P}\left(\theta_0 \notin \left[\min_{1 \leq j \leq B} \hat{\theta}_j,\ \max_{1 \leq j \leq B} \hat{\theta}_j \right] \right) &\leq P(B; \Delta) \\ &= \left(\frac{1}{2} - \Delta \right)^B + \left(\frac{1}{2} + \Delta \right)^B. \end{aligned}$$

Adaptive HulC

$$\Delta := \max_{j \geq 1} \left(\frac{1}{2} - \min \left\{ \mathbb{P}(\hat{\theta}_j \leq \theta_0),\, \mathbb{P}(\hat{\theta}_j \geq \theta_0) \right\} \right)_+ ,$$

can be estimated using subsampling:

b is the subsample size.

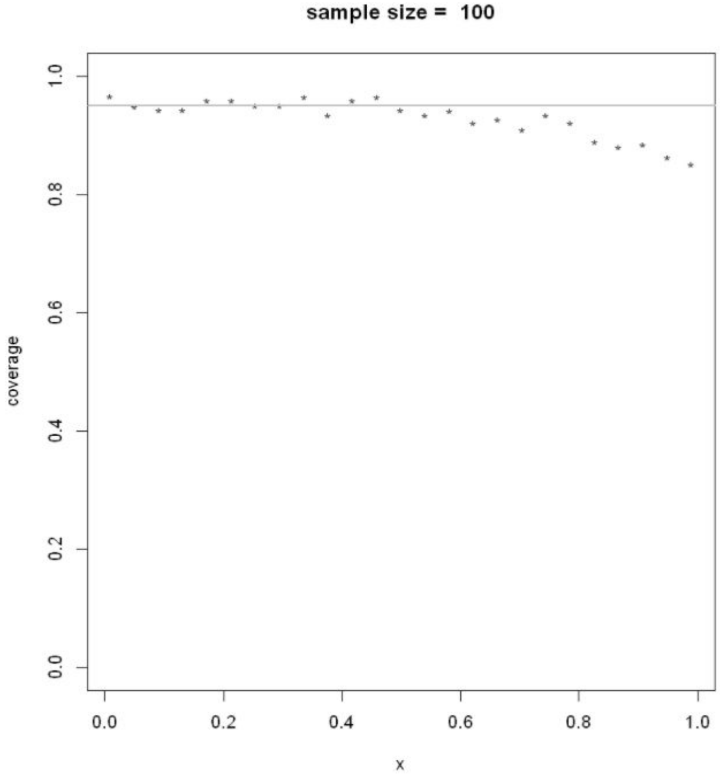
$$\widehat{\Delta} := \left| \frac{1}{2} - \frac{1}{K_n} \sum_{j=1}^{K_n} \mathbf{1}\{\hat{\theta}_j^{(b)} \leq \hat{\theta}\} \right| .$$

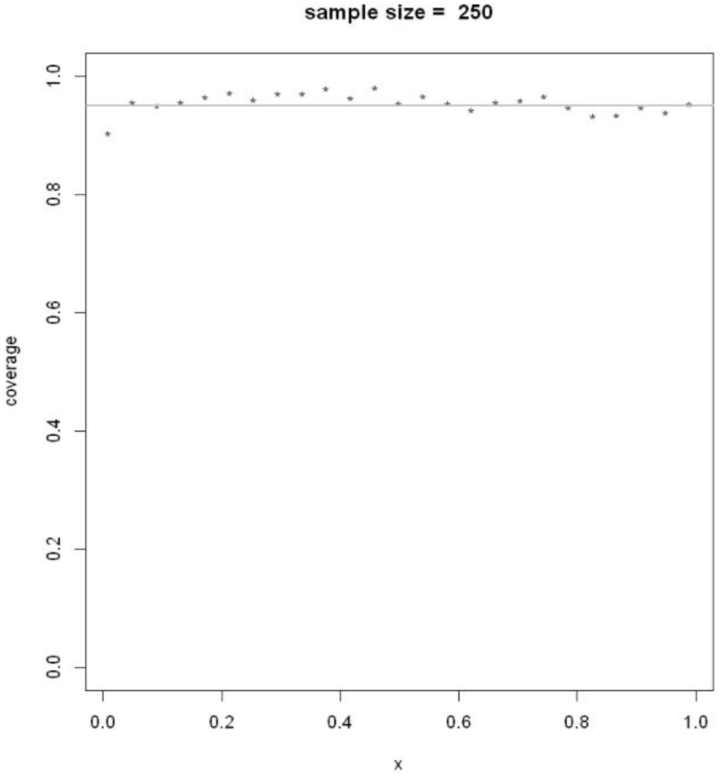
This does not require knowing the rate of convergence of the estimator, while the traditional application of subsampling requires such knowledge.

Monotone Regression: LSE

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

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





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.
- HulC and adaptive HulC together can be used in many settings including nonparametric regression problems.

Thanks for your attention!!