Adaptive Inference in Irregular M-estimation

Kenta Takatsu, Carnegie Mellon University
Based on joint works with Arun Kumar Kuchibhotla
IISA Student Paper Competition, June 2025

Bridging Root-n and Non-standard Asymptotics: Adaptive Inference in M-Estimation

Kenta Takatsu and Arun Kumar Kuchibhotla

Department of Statistics and Data Science, Carnegie Mellon University

Abstract

This manuscript studies a general approach to construct confidence sets for the solution of population-level optimization, commonly referred to as M-estimation. Statistical inference for M-estimation poses significant challenges due to the non-standard limiting behaviors of the corresponding estimator, which arise in settings with increasing dimension of parameters, non-smooth objectives, or constraints. We propose a simple and unified method that guarantees validity in both regular and irregular cases. Moreover, we provide a comprehensive width analysis of the proposed confidence set, showing that the convergence rate of the diameter is adaptive to the unknown degree of instance-specific regularity. We apply the proposed method to several high-dimensional and irregular statistical problems.

Keywords— Honest inference, Adaptive inference, Irregular M-estimation, Non-standard asymptotics, Extremum estimators.

arXiv:2501.07772

Given observations $\{X_i\}_{i=1}^n$ from unknown distribution $P \in \mathscr{P}$, we are interested in some “summary" of P.

We consider the summary as minimizer of expected loss function:

$$P \mapsto \theta_P := \arg\min_{\theta \in \Theta} \mathbb{E}_P[m(X;\theta)].$$

This is called M-estimation

Given observations $\{X_i\}_{i=1}^n$ from unknown distribution $P \in \mathscr{P}$, we are interested in some "summary" of P.

We consider the summary as minimizer of expected loss function:

$$P \mapsto \theta_P := \arg\min_{\theta \in \Theta} \mathbb{E}_P[m(X;\theta)].$$

Mean / Median	MLE	Regression fn.
Classification	Model selection	Discrete choice

The parameter space Θ can be **nonparametric/high-dimensional**, **constrained** (shape/sparsity), or **discontinuous**.

Goal: Construct a confidence set $\mathrm{CI}_{n,\alpha}$ for $\alpha \in [0,1]$ such that

$$\sup_{P \in \mathscr{P}} \mathbb{P}(\theta_P \notin \mathrm{CI}_{n,\alpha}) \leq \alpha.$$

"Traditional" approach

Goal: Construct a confidence set $\mathrm{CI}_{n,\alpha}$ for $\alpha \in [0,1]$ such that

$$\sup_{P \in \mathscr{P}} \mathbb{P}(\theta_P \notin \mathrm{CI}_{n,\alpha}) \leq \alpha.$$

"Traditional" approach

1. Construct an estimator $\widehat{\theta}$ of θ_P.

2. Establish convergence in distribution:

$$r_n(\widehat{\theta} - \theta_P) \xrightarrow{d} G_P \qquad (1)$$

3. Invert this expression (1):

$$\mathrm{CI}_{n,\alpha} := [\widehat{\theta} - r_n^{-1}\hat{q}_{1-\alpha/2}, \widehat{\theta} + r_n^{-1}\hat{q}_{\alpha/2}]$$

Goal: Construct a confidence set $\mathrm{CI}_{n,\alpha}$ for $\alpha \in [0,1]$ such that

$$\sup_{P \in \mathcal{P}} \mathbb{P}(\theta_P \notin \mathrm{CI}_{n,\alpha}) \leq \alpha.$$

"Traditional" approach

1. Construct an estimator $\widehat{\theta}$ of θ_P.

2. Establish convergence in distribution:

$$r_n(\widehat{\theta} - \theta_P) \xrightarrow{d} G_P \qquad (1)$$

3. Invert this expression (1):

$$\mathrm{CI}_{n,\alpha} := [\widehat{\theta} - r_n^{-1}\hat{q}_{1-\alpha/2}, \widehat{\theta} + r_n^{-1}\hat{q}_{\alpha/2}]$$

Example

$$n^{1/2}(\widehat{\theta} - \theta_P) \xrightarrow{d} N(0,\sigma_P^2)$$

$$\mathrm{CI}_{n,\alpha} := [\widehat{\theta} \pm z_{\alpha/2} n^{-1/2} \widehat{\sigma}_P]$$

Failure of Wald and Resampling Inference

The problem is $r_n(\widehat{\theta} - \theta_P) \xrightarrow{d} G_P$

[Scheffé and Tukey, (1945); Smirnov, (1952)]

Failure of Wald and Resampling Inference

The problem is $r_n(\widehat{\theta} - \theta_P) \xrightarrow{d} G_P$

Example

Suppose θ_P is median and $\widehat{\theta}$ is sample median:

[Scheffé and Tukey, (1945); Smirnov, (1952)]

Failure of Wald and Resampling Inference

The problem is $r_n(\widehat{\theta} - \theta_P) \xrightarrow{d} G_P$

Example Suppose θ_P is median and $\widehat{\theta}$ is sample median:

Under regularity condition, $r_n = n^{1/2}$ and the limiting distribution is Gaussian.

Regular

[Scheffé and Tukey, (1945); Smirnov, (1952)]

Failure of Wald and Resampling Inference

The problem is $r_n(\widehat{\theta} - \theta_P) \xrightarrow{d} G_P$

Example Suppose θ_P is median and $\widehat{\theta}$ is sample median:

Under regularity condition, $r_n = n^{1/2}$ and the limiting distribution is Gaussian.

Regular

↔

Otherwise, $r_n = n^{1/(2\beta)}$ and the limiting distribution is non-Gaussian, both depend on an unknown parameter β.

Irregular

[Scheffé and Tukey, (1945); Smirnov, (1952)]

Failure of Wald and Resampling Inference

The problem is $r_n(\widehat{\theta} - \theta_P) \xrightarrow{d} G_P$

Most commonly used targets θ_P are M-estimands, defined by

$$\theta_P := \arg\min_{\theta \in \Theta} \mathbb{E}_P[m(X; \theta)]$$

Failure of traditional inference is also observed, for instance, when

the parameter space Θ is **high-dimensional**;

the parameter space Θ is **constrained;**

the minimizer θ_P is near/on the **boundary** of Θ;

the mapping $\theta \mapsto \mathbb{E}_P[m(X; \theta)]$ is **non-smooth** near θ_P, and so on.

Statistical inference for irregular M-estimation is an ongoing challenge.

- Subsampling/Bootstrap typically fail for these problems.
- We don't always know the rate of convergence or limiting distribution of the standard estimator.

Statistical inference for irregular M-estimation is an ongoing challenge.

Subsampling/Bootstrap typically fail for these problems.

We don't always know the rate of convergence or limiting distribution of the standard estimator.

Regardless, we construct a confidence set $\mathrm{CI}_{n,\alpha}$ that

(1) remains valid without the knowledge of the regularity;

(2) converges adaptively at a rate depending on the regularity.

This is adaptive inference

Proposed Procedure

***T.** and Kuchibhotla, A. K. (2025)*

Recall $\theta_P := \arg\min_{\theta \in \Theta} \mathbb{M}(\theta)$ and $\mathbb{M}(\theta) = \mathbb{E}_P[m(X; \theta)]$

We employ sample-splitting

Given $2n$ samples, we construct *any* estimator $\widehat{\theta}$ using the first half.

On the second half, we perform the following:

Proposed Procedure

T. and Kuchibhotla, A. K. (2025)

Recall $\theta_P := \arg\min_{\theta \in \Theta} \mathbb{M}(\theta)$ and $\mathbb{M}(\theta) = \mathbb{E}_P[m(X; \theta)]$

Given $2n$ samples, we construct *any* estimator $\widehat{\theta}$ using the first half.

On the second half, we perform the following:

1. For each $\theta \in \Theta$, compute the difference of empirical losses:

$$\widehat{\mathbb{M}}(\theta) - \widehat{\mathbb{M}}(\widehat{\theta}) = n^{-1} \sum_{i=n+1}^{2n} m(X_i; \theta) - m(X_i; \widehat{\theta}).$$

2. Report the confidence set: $\mathrm{CI}_{n,\alpha} := \left\{ \theta \in \Theta \, : \, \widehat{\mathbb{M}}(\theta) - \widehat{\mathbb{M}}(\widehat{\theta}) \leq z_\alpha n^{-1/2} \widehat{\sigma}_\theta \right\}$

   where $n^{-1/2} \widehat{\sigma}_\theta$ is an estimate of the standard deviation of $\widehat{\mathbb{M}}(\theta) - \widehat{\mathbb{M}}(\widehat{\theta})$.

Proposed Procedure

T. and Kuchibhotla, A. K. (2025)

Recall $\theta_P := \arg\min_{\theta\in\Theta} \mathbb{M}(\theta)$ and $\mathbb{M}(\theta) = \mathbb{E}_P[m(X;\theta)]$

Given $2n$ samples, we construct *any* estimator $\widehat{\theta}$ using the first half.

On the second half, we perform the following:

1. For each $\theta \in \Theta$, compute the difference of empirical losses:

$$\widehat{\mathbb{M}}(\theta) - \widehat{\mathbb{M}}(\widehat{\theta}) = n^{-1} \sum_{i=n+1}^{2n} m(X_i;\theta) - m(X_i;\widehat{\theta}).$$

2. Report the confidence set: $\mathrm{CI}_{n,\alpha} := \left\{ \theta \in \Theta \,:\, \widehat{\mathbb{M}}(\theta) - \widehat{\mathbb{M}}(\widehat{\theta}) \le z_\alpha n^{-1/2} \widehat{\sigma}_\theta \right\}$

   where $n^{-1/2}\widehat{\sigma}_\theta$ is an estimate of the standard deviation of $\widehat{\mathbb{M}}(\theta) - \widehat{\mathbb{M}}(\widehat{\theta})$.

Intuition

By definition θ_P lies in the set $\left\{ \theta \in \Theta \,:\, \mathbb{M}(\theta) - \mathbb{M}(\widehat{\theta}) \le 0 \right\}$. A key observation is that the risk of an irregular estimator may be well-behaving.

A Brief History

Inverting the risk of an irregular estimator has a long history.

A Brief History

Inverting the risk of an irregular estimator has a long history.

Stein mentioned the idea in passing

1981

[Stein, 1981]

A Brief History

Inverting the risk of an irregular estimator has a long history.

The inversion based on CLT appeared in the late 1990s

1981

1996~1998

[Beran, 1996; Beran and Dümbgen, 1998]

A Brief History

Inverting the risk of an irregular estimator has a long history.

[Robins and van der Vaart, 2006]

A Brief History

Inverting the risk of an irregular estimator has a long history.

Many recent works use this idea for irregular inference

1981

1996~1998

2006

2018 ~

[Chakravarti et al. (2019); Kim and Ramdas (2024); Park et al. (2025+); Takatsu and Kuchibhotla (2025+)]

A Brief History

Inverting the risk of an irregular estimator has a long history.

[Shapiro (1989); Geyer (1994); Pflug (1991, 1995, 2003); Vogel (2008)]

A Brief History

Inverting the risk of an irregular estimator has a long history.

Our contribution in [**T.** and Kuchibhotla, A. K. (2025)] lies in analyzing the validity and width properties for general M-estimation problems

Properties of the Confidence Set

T. *and Kuchibhotla, A. K. (2025)*

Properties of the Confidence Set

***T.** and Kuchibhotla, A. K. (2025)*

Validity

We do not need to know the rate of convergence or the limiting distribution of the M-estimator for our method.

Validity holds even when θ_P is not unique.

By sample-splitting, validity holds regardless of the dimension/complexity of Θ or the choice of $\widehat{\theta}$.

Properties of the Confidence Set

T. and Kuchibhotla, A. K. (2025)

Validity

We do not need to know the rate of convergence or the limiting distribution of the M-estimator for our method.

Validity holds even when θ_P is not unique.

By sample-splitting, validity holds regardless of the dimension/complexity of Θ or the choice of $\widehat{\theta}$.

Size of the CI

We assume θ_P is unique for the analysis.

The diameter converges at an adaptive rate, depending on the geometry of the problem.

The convergence rate also depends on the dimension/complexity of Θ and the choice of $\widehat{\theta}$.

Properties of the Confidence Set

T. and Kuchibhotla, A. K. (2025)

Validity

We do not need to know the rate of convergence or the limiting distribution of the M-estimator for our method.

Validity holds even when θ_P is not unique.

By sample-splitting, validity holds regardless of the dimension/complexity of Θ or the choice of $\widehat{\theta}$.

Size of the CI

We assume θ_P is unique for the analysis.

The diameter converges at an adaptive rate, depending on the geometry of the problem.

The convergence rate also depends on the dimension/complexity of Θ and the choice of $\widehat{\theta}$.

Application

High-dimensional MLEs; High-dimensional regression; Manski's maximum score estimator; Quantile; Argmin.

Validity Condition

***T.** and Kuchibhotla, A. K. (2025)*

Recall $\mathbb{M}(\theta) = \mathbb{E}_P[m(X;\theta)]$. Define $\xi_{P,i} := m(X_i;\theta_P) - m(X_i;\widehat{\theta})$ and $\sigma_P^2 := \mathrm{Var}[\xi_P \mid \widehat{\theta}]$.

Theorem

For any $n \geq 1$,

$$\mathbb{P}_P(\theta_P \notin \mathrm{CI}_{n,\alpha} \mid \widehat{\theta}) \leq \min\left\{ \frac{\sigma_P^2}{n \,|\, \mathbb{M}(\theta_P) - \mathbb{M}(\widehat{\theta}) \,|^2},\ \alpha + \mathbb{E}_P\left[\frac{|\,\xi_P - \mathbb{E}_P[\xi_P]\,|^3}{n^{1/2}\sigma_P^3} \,\middle|\, \widehat{\theta} \right] \right\}.$$

We verify that the RHS converges to zero uniformly over large collection of distributions $\mathscr{P}$ in regular problems (QMD) and also several irregular problems, including Manski model, quantile estimation and constrained problems.

Crucially, the RHS does not depend on the dimension of Θ

Convergence Rates

T. and Kuchibhotla, A. K. (2025)

For all $\theta \in \Theta$,

Curvature $\mathbb{E}_P[m(X;\theta) - m(X;\theta_P)] \gtrsim \|\theta - \theta_P\|^{1+\beta}$ for some $\beta \geq 0$.

Variance $\mathrm{Var}_P[m(X;\theta) - m(X;\theta_P)] \lesssim \|\theta - \theta_P\|^{2\eta}$ for some $\eta < 1 + \beta$.

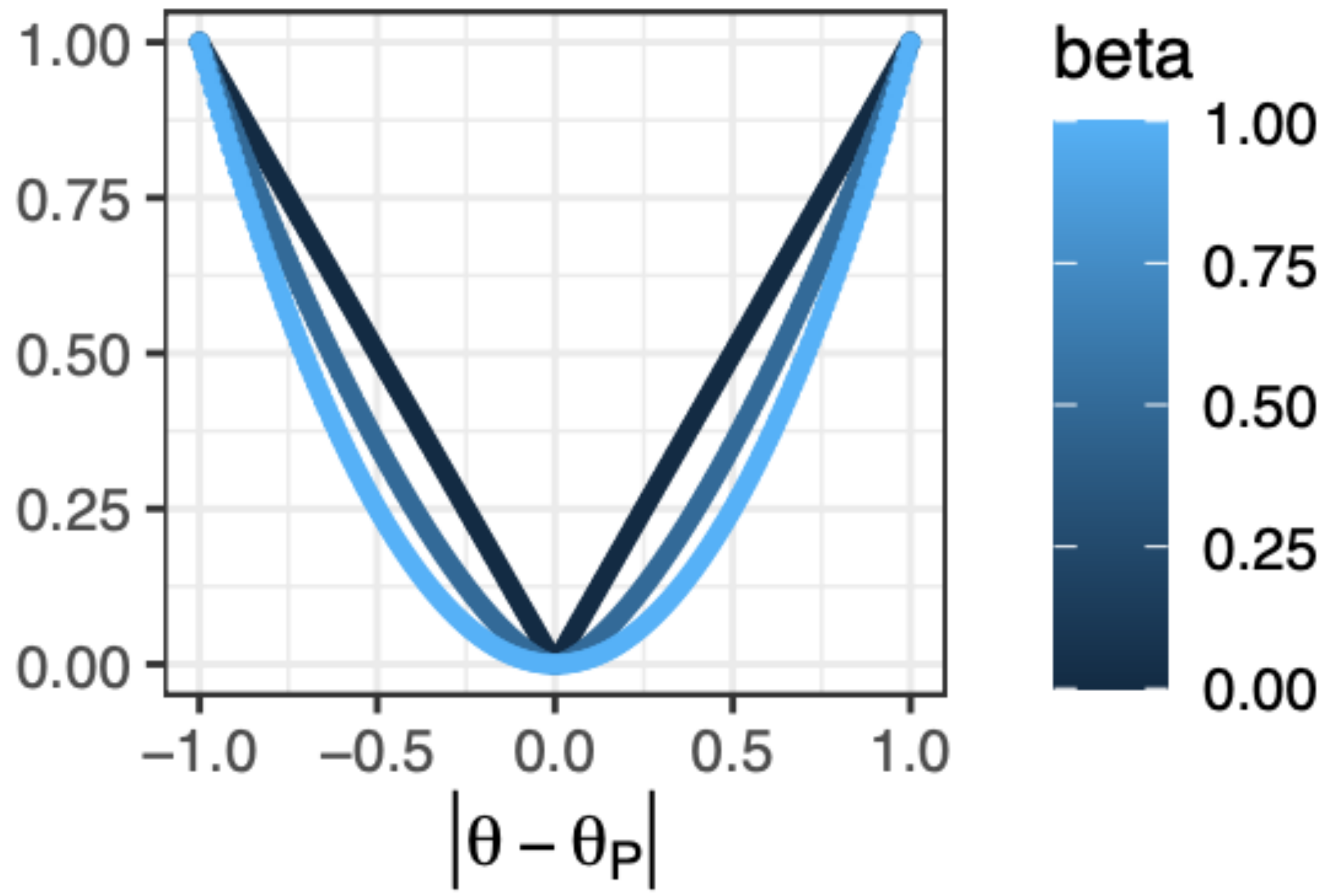


Illustration of curvature

Convergence Rates

T. and Kuchibhotla, A. K. (2025)

For all $\theta \in \Theta$,

Curvature $\mathbb{E}_P[m(X;\theta) - m(X;\theta_P)] \gtrsim \|\theta - \theta_P\|^{1+\beta}$ for some $\beta \geq 0$.

Variance $\mathrm{Var}_P[m(X;\theta) - m(X;\theta_P)] \lesssim \|\theta - \theta_P\|^{2\eta}$ for some $\eta < 1 + \beta$.

Theorem (Informal)

The diameter of the confidence set satisfies

$$\mathrm{Diam}_{\|\cdot\|}(\mathrm{CI}_{n,\alpha}) = O_P(n^{-1/(2+2\beta-2\eta)} + r_n^{1/(1+\beta)} + s_n^{1/(1+\beta)}).$$

When $\beta = \eta = 1$, we get a parametric rate

When $\beta = 1$ and $\eta = 1/2$, we get a cube-root rate

* $\mathrm{Diam}_{\|\cdot\|}(A) := \sup\{\|a - b\| \ : \ a, b \in A\}$.

Convergence Rates

T. and Kuchibhotla, A. K. (2025)

For all $\theta \in \Theta$,

Curvature $\mathbb{E}_P[m(X;\theta) - m(X;\theta_P)] \gtrsim \|\theta - \theta_P\|^{1+\beta}$ for some $\beta \geq 0$.

Variance $\mathrm{Var}_P[m(X;\theta) - m(X;\theta_P)] \lesssim \|\theta - \theta_P\|^{2\eta}$ for some $\eta < 1 + \beta$.

Theorem (Informal)

The diameter of the confidence set satisfies

$$\mathrm{Diam}_{\|\cdot\|}(\mathrm{CI}_{n,\alpha}) = O_P(n^{-1/(2+2\beta-2\eta)} + r_n^{1/(1+\beta)} + s_n^{1/(1+\beta)}).$$

The confidence set converges adaptively to unknown β and η.

When $\beta = \eta = 1$, we get a parametric rate

When $\beta = 1$ and $\eta = 1/2$, we get a cube-root rate

* $\mathrm{Diam}_{\|\cdot\|}(A) := \sup\{\|a - b\| \,:\, a, b \in A\}$.

Convergence Rates

T. and Kuchibhotla, A. K. (2025)

For all $\theta \in \Theta$,

Curvature $\mathbb{E}_P[m(X;\theta) - m(X;\theta_P)] \gtrsim \|\theta - \theta_P\|^{1+\beta}$ for some $\beta \geq 0$.

Variance $\mathrm{Var}_P[m(X;\theta) - m(X;\theta_P)] \lesssim \|\theta - \theta_P\|^{2\eta}$ for some $\eta < 1 + \beta$.

Theorem (Informal)

The diameter of the confidence set satisfies

$$\mathrm{Diam}_{\|\cdot\|}(\mathrm{CI}_{n,\alpha}) = O_P(n^{-1/(2+2\beta-2\eta)} + r_n^{1/(1+\beta)} + s_n^{1/(1+\beta)}).$$

r_n depends on the complexity of Θ, and the moments of the local envelope

$$\sup_{\|\theta-\theta_P\|<\delta} |m(X;\theta) - m(X;\theta_P)|$$

When $\beta = \eta = 1$, we get a parametric rate

When $\beta = 1$ and $\eta = 1/2$, we get a cube-root rate

* $\mathrm{Diam}_{\|\cdot\|}(A) := \sup\{\|a - b\| \ : \ a, b \in A\}$.

Convergence Rates

T. and Kuchibhotla, A. K. (2025)

For all $\theta \in \Theta$,

Curvature $\mathbb{E}_P[m(X;\theta) - m(X;\theta_P)] \gtrsim \|\theta - \theta_P\|^{1+\beta}$ for some $\beta \geq 0$.

Variance $\mathrm{Var}_P[m(X;\theta) - m(X;\theta_P)] \lesssim \|\theta - \theta_P\|^{2\eta}$ for some $\eta < 1 + \beta$.

Theorem (Informal)

The diameter of the confidence set satisfies

$$\mathrm{Diam}_{\|\cdot\|}(\mathrm{CI}_{n,\alpha}) = O_P(n^{-1/(2+2\beta-2\eta)} + r_n^{1/(1+\beta)} + s_n^{1/(1+\beta)}).$$

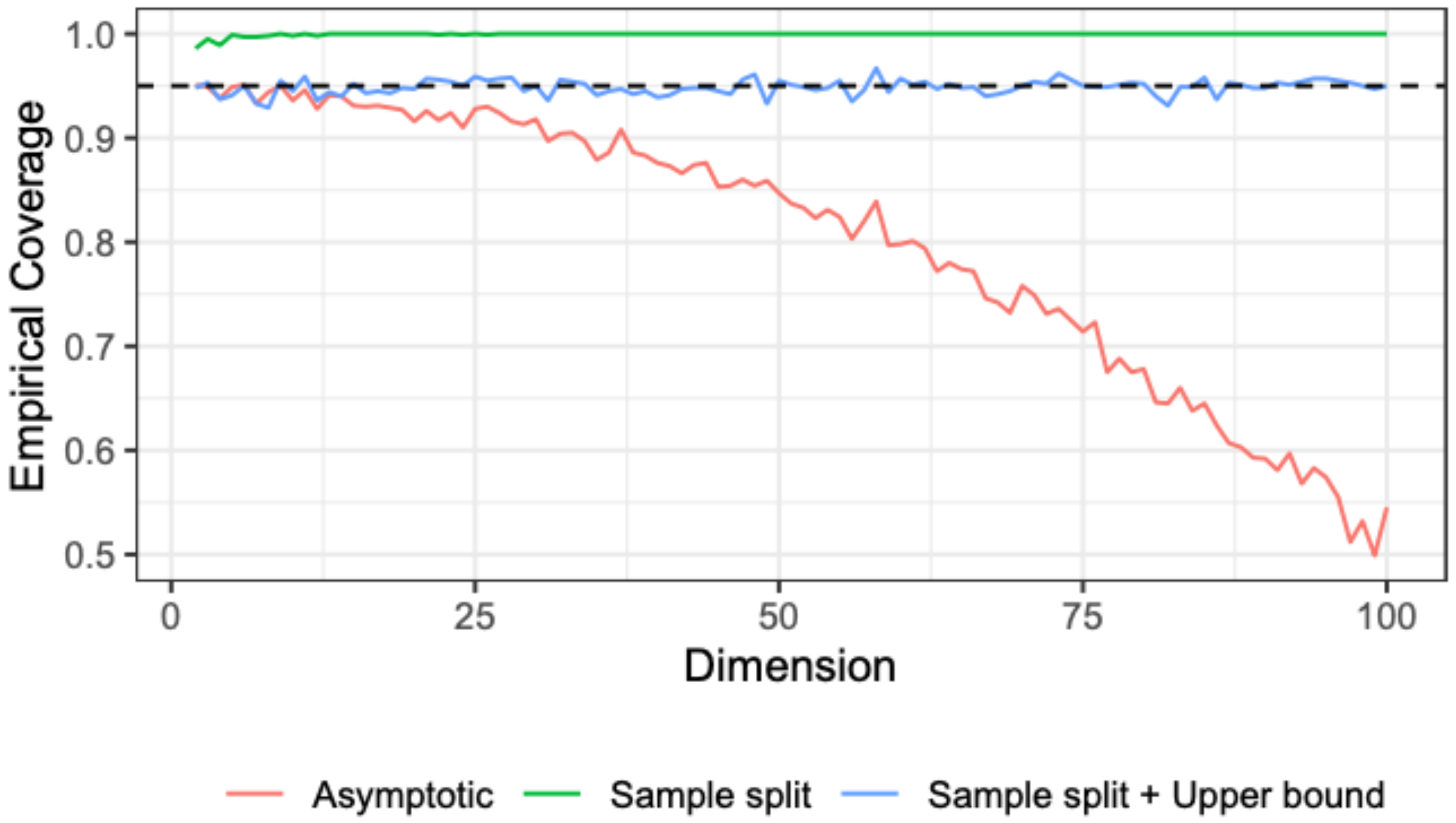
s_n depends on the convergence rate of the initial estimator

When $\beta = \eta = 1$, we get a parametric rate
When $\beta = 1$ and $\eta = 1/2$, we get a cube-root rate

* $\mathrm{Diam}_{\|\cdot\|}(A) := \sup\{\|a - b\| \ : \ a, b \in A\}$.

Example 1 (Mean Inference)

Consider high-dimensional mean inference where $\theta_P := \arg\min_{\theta \in \Theta} \mathbb{E}_P \|X - \theta\|_2^2$.



D-dimensional mean inference from N = 500.

Example 2 (Manski's Discrete Choice Model)

Consider an IID observation $(Y_1, X_1), \ldots, (Y_{2n}, X_{2n}) \in \{-1,1\} \times \mathbb{R}^d$ generated from Manski's model:

$$Y_i := \operatorname{sgn}(\theta_P^\top X_i + \varepsilon_i) \text{ where } \operatorname{Med}(\varepsilon_i \mid X_i) = 0$$

It has been shown that $\theta_P = \arg\max_{\theta \in \mathbb{S}^{d-1}} \mathbb{E}_P[Y \operatorname{sgn}(\theta_P^\top X)]$.

* $\operatorname{sgn}(t) = 2\mathbf{1}\{t \geq 0\} - 1.$

[Manski, (1985)]

Example 2 (Manski's Discrete Choice Model)

Recall $\theta_P = \arg\max_{\theta \in \mathbb{S}^{d-1}} \mathbb{E}_P[Y \operatorname{sgn}(\theta_P^\top X)]$.

A1 Set $\eta(x) = \mathbb{P}_X(Y = 1 \mid X = x)$. Assume $\mathbb{P}_X(|\eta(X) - 1/2| > t) \lesssim t^{1/\beta}$ for all $t \geq 0$.

A2 Assume $\|\theta - \theta_P\|_2 \lesssim \mathbb{P}_X(\operatorname{sgn}(\theta^\top X) \neq \operatorname{sgn}(\theta_P^\top X))$ for all $\theta \in \mathbb{S}^{d-1}$.

Theorem (Informal)

Under the assumptions **A1** and **A2**:

$$\operatorname{Diam}_{\|\cdot\|}(\mathrm{CI}_{n,\alpha}) = O_P\left(\left(\frac{d \log(d/n)}{n}\right)^{1/\beta} + s_n^{1/(1+2\beta)}\right).$$

This matches the known minimax estimation rate.

[Mukherjee et al. (2019, 2021)]

Summary

Risk inversion and sample-splitting provide a general confidence set for (irregular) M-estimation.

The confidence set is valid under very weak assumptions. For instance, we do not need to know the rate of convergence or the limiting distribution of the M-estimator, and the validity of this method is dimension-free.

The width of the set converges at an adaptive rate, depending on the (unknown) geometry of the problems, such as the curvature.

Thank You

Bridging Root-n and Non-standard Asymptotics: Adaptive Inference in M-Estimation

Kenta Takatsu and Arun Kumar Kuchibhotla

Department of Statistics and Data Science, Carnegie Mellon University

Abstract

This manuscript studies a general approach to construct confidence sets for the solution of population-level optimization, commonly referred to as M-estimation. Statistical inference for M-estimation poses significant challenges due to the non-standard limiting behaviors of the corresponding estimator, which arise in settings with increasing dimension of parameters, non-smooth objectives, or constraints. We propose a simple and unified method that guarantees validity in both regular and irregular cases. Moreover, we provide a comprehensive width analysis of the proposed confidence set, showing that the convergence rate of the diameter is adaptive to the unknown degree of instance-specific regularity. We apply the proposed method to several high-dimensional and irregular statistical problems.

Keywords— Honest inference, Adaptive inference, Irregular M-estimation, Non-standard asymptotics, Extremum estimators.

arXiv:2501.07772

Appendix

Bridging Root-n and Non-standard Asymptotics: Adaptive Inference in M-Estimation

Kenta Takatsu and Arun Kumar Kuchibhotla

Department of Statistics and Data Science, Carnegie Mellon University

Abstract

This manuscript studies a general approach to construct confidence sets for the solution of population-level optimization, commonly referred to as M-estimation. Statistical inference for M-estimation poses significant challenges due to the non-standard limiting behaviors of the corresponding estimator, which arise in settings with increasing dimension of parameters, non-smooth objectives, or constraints. We propose a simple and unified method that guarantees validity in both regular and irregular cases. Moreover, we provide a comprehensive width analysis of the proposed confidence set, showing that the convergence rate of the diameter is adaptive to the unknown degree of instance-specific regularity. We apply the proposed method to several high-dimensional and irregular statistical problems.

Keywords— Honest inference, Adaptive inference, Irregular M-estimation, Non-standard asymptotics, Extremum estimators.

arXiv:2501.07772

On the Precise Asymptotics of Universal Inference

Kenta Takatsu

Department of Statistics and Data Science, Carnegie Mellon University

Abstract

In statistical inference, confidence set procedures are typically evaluated based on their validity and width properties. Even when procedures achieve rate-optimal widths, confidence sets can still be excessively wide in practice due to elusive constants, leading to extreme conservativeness, where the empirical coverage probability of nominal $1-\alpha$ level confidence sets approaches one. This manuscript studies this gap between validity and conservativeness, using *universal inference* (Wasserman et al., 2020) with a regular parametric model under model misspecification as a running example. We identify the source of asymptotic conservativeness and propose a general remedy based on *studentization* and *bias correction*. The resulting method attains exact asymptotic coverage at the nominal $1-\alpha$ level, even under model misspecification, provided that the product of the estimation errors of two unknowns is negligible, exhibiting an intriguing resemblance to double robustness in semiparametric theory.

Keywords— Universal Inference, Central Limit Theorem, Berry-Esseen Bound, Model Misspecification, Studentization, Double Robustness

arXiv:2503.14717

Key Idea: Risk Inversion

Observe that θ_P is a minimizer and $\mathbb{E}_P[m(X;\theta_P)] - \mathbb{E}_P[m(X;\widehat{\theta}) \mid \widehat{\theta}] \leq 0$ for *any* $\widehat{\theta} \in \Theta$.

Key Idea: Risk Inversion

Observe that θ_P is a minimizer and $\mathbb{E}_P[m(X;\theta_P)] - \mathbb{E}_P[m(X;\hat{\theta}) \mid \hat{\theta}] \leq 0$ for *any* $\hat{\theta} \in \Theta$.

A non-actionable but valid confidence set is $\left\{\theta \in \Theta \, : \, \mathbb{E}_P[m(X;\theta) - m(X;\hat{\theta}) \mid \hat{\theta}] \leq 0\right\}$.

Key Idea: Risk Inversion

Observe that θ_P is a minimizer and $\mathbb{E}_P[m(X;\theta_P)] - \mathbb{E}_P[m(X;\hat{\theta}) \mid \hat{\theta}] \le 0$ for *any* $\hat{\theta} \in \Theta$.

A non-actionable but valid confidence set is $\left\{\theta \in \Theta \,:\, \mathbb{E}_P[m(X;\theta) - m(X;\hat{\theta}) \mid \hat{\theta}] \le 0\right\}$.

The proposed confidence set is $\left\{\theta \in \Theta \,:\, n^{-1}\sum [m(X_i;\theta) - m(X_i;\hat{\theta})] \le \gamma_{n,\alpha}\right\}$ where $\gamma_{n,\alpha} \to 0$ is an appropriate cutoff to guarantee validity.

Key Idea: Risk Inversion

Observe that θ_P is a minimizer and $\mathbb{E}_P[m(X;\theta_P)] - \mathbb{E}_P[m(X;\hat{\theta}) \mid \hat{\theta}] \leq 0$ for *any* $\hat{\theta} \in \Theta$.

A non-actionable but valid confidence set is $\left\{\theta \in \Theta \,:\, \mathbb{E}_P[m(X;\theta) - m(X;\hat{\theta}) \mid \hat{\theta}] \leq 0\right\}$.

The proposed confidence set is $\left\{\theta \in \Theta \,:\, n^{-1} \sum [m(X_i;\theta) - m(X_i;\hat{\theta})] \leq \gamma_{n,\alpha}\right\}$ where $\gamma_{n,\alpha} \to 0$ is an appropriate cutoff to guarantee validity.

From earlier, we define $\xi_i := m(X_i;\theta) - m(X_i;\hat{\theta})$, and we can use the central limit theorem (CLT) for the t-statistics of $\{\xi_i\}$ to obtain $\gamma_{n,\alpha}$.

Key Idea: Risk Inversion

Observe that θ_P is a minimizer and $\mathbb{E}_P[m(X;\theta_P)] - \mathbb{E}_P[m(X;\widehat{\theta}) \mid \widehat{\theta}] \le 0$ for *any* $\widehat{\theta} \in \Theta$.

A non-actionable but valid confidence set is $\left\{\theta \in \Theta \,:\, \mathbb{E}_P[m(X;\theta) - m(X;\widehat{\theta}) \mid \widehat{\theta}] \le 0\right\}$.

The proposed confidence set is $\left\{\theta \in \Theta \,:\, n^{-1}\sum [m(X_i;\theta) - m(X_i;\widehat{\theta})] \le \gamma_{n,\alpha}\right\}$ where $\gamma_{n,\alpha} \to 0$ is an appropriate cutoff to guarantee validity.

From earlier, we define $\xi_i := m(X_i;\theta) - m(X_i;\widehat{\theta})$, and we can use the central limit theorem (CLT) for the t-statistics of $\{\xi_i\}$ to obtain $\gamma_{n,\alpha}$.

Similar ideas exist in the literature [Beran and Dümbgen, 1996; Robins and van der Vaart, 2006; Vogel, 2008].
Our contribution is in new theoretical analysis and statistical applications.

Validity Condition

T. and Kuchibhotla, A. K. (2025)

Define $\xi_{P,i} := m(X_i; \theta_P) - m(X_i; \hat{\theta})$. Define sample mean and variance as $\bar{\xi}_P$ and $\hat{\sigma}_P^2$.

Denote the (conditional) Kolmogorov-Smirnov distance by

$$\Delta_{n,P} := \sup_{t \in \mathbb{R}} \left| \mathbb{P}_P \left(\frac{n^{1/2}(\bar{\xi}_P - \mathbb{E}[\bar{\xi}_P])}{\hat{\sigma}_P} \leq t \,|\, \hat{\theta} \right) - \Phi(t) \right|$$

where $\Phi(t)$ is the CDF of the standard Normal.

Validity Condition

***T.** and Kuchibhotla, A. K. (2025)*

Define $\xi_{P,i} := m(X_i; \theta_P) - m(X_i; \hat{\theta})$. Define sample mean and variance as $\bar{\xi}_P$ and $\hat{\sigma}_P^2$.

Denote the (conditional) Kolmogorov-Smirnov distance by

$$\Delta_{n,P} := \sup_{t \in \mathbb{R}} \left| \mathbb{P}_P \left(\frac{n^{1/2}(\bar{\xi}_P - \mathbb{E}[\bar{\xi}_P])}{\hat{\sigma}_P} \leq t \,|\, \hat{\theta} \right) - \Phi(t) \right|$$

where $\Phi(t)$ is the CDF of the standard Normal.

Theorem

For any $n \geq 1$, it holds $\inf_{P \in \mathscr{P}} \mathbb{P}_P(\theta_P \in \mathrm{CI}_{n,\alpha}) \geq 1 - \alpha - \sup_{P \in \mathscr{P}} \mathbb{E}_P[\Delta_{n,P}]$.

Validity Condition

***T.** and Kuchibhotla, A. K. (2025)*

Define $\xi_{P,i} := m(X_i; \theta_P) - m(X_i; \widehat{\theta})$ and $\sigma_P^2 := \mathrm{Var}[\xi_P \mid \widehat{\theta}]$.

For any $n \geq 1$,

$$\Delta_{n,P} \leq \min\left\{1, \mathbb{E}_P\left[\frac{|\xi_P - \mathbb{E}_P[\xi_P]|^2}{\sigma_P^2} \min\left\{1, \frac{|\xi_P - \mathbb{E}_P[\xi_P]|}{n^{1/2}\sigma_P}\right\} \,\middle|\, \widehat{\theta}\right]\right\}.$$

Berry-Esseen bound for t-statistics (Katz, 1963; Bentkus et al., 1996)

Validity Condition

T. and Kuchibhotla, A. K. (2025)

Define $\xi_{P,i} := m(X_i; \theta_P) - m(X_i; \widehat{\theta})$ and $\sigma_P^2 := \mathrm{Var}[\xi_P \mid \widehat{\theta}]$.

For any $n \geq 1$,

$$\Delta_{n,P} \leq \min\left\{1, \mathbb{E}_P\left[\frac{|\xi_P - \mathbb{E}_P[\xi_P]|^2}{\sigma_P^2} \min\left\{1, \frac{|\xi_P - \mathbb{E}_P[\xi_P]|}{n^{1/2}\sigma_P}\right\} \,\middle|\, \widehat{\theta}\right]\right\}.$$

Berry-Esseen bound for t-statistics (Katz, 1963; Bentkus et al., 1996)

We provide conditions on $\{\xi_{P,i}\}$ under which $\Delta_{n,P} = o_P(1)$ as $n \to \infty$ uniformly over $\mathscr{P}$.

This holds under mild assumptions on P, including the cases traditionally considered "irregular".

Crucially, this expression does not depend on the dimension of Θ as $\xi_{P,i} \in \mathbb{R}$.

Conservativeness

T. (2025)

We have built a confidence set $\mathrm{CI}_{n,\alpha}$ such that

(1) remains valid without the knowledge of the regularity;

(2) shrinks adaptively at a rate depending on the regularity.

Conservativeness

***T.** (2025)*

We have built a confidence set $\mathrm{CI}_{n,\alpha}$ such that

(1) remains valid without the knowledge of the regularity;

(2) shrinks adaptively at a rate depending on the regularity.

Even when (1) and (2) hold, the confidence set can be overly large, in other words, too **conservative**.

Question:

Is it $\inf_{P \in \mathscr{P}} \mathbb{P}(\theta_P \in \mathrm{CI}_{n,\alpha}) \approx 1 - \alpha$

or $\inf_{P \in \mathscr{P}} \mathbb{P}(\theta_P \in \mathrm{CI}_{n,\alpha}) \approx 1$?

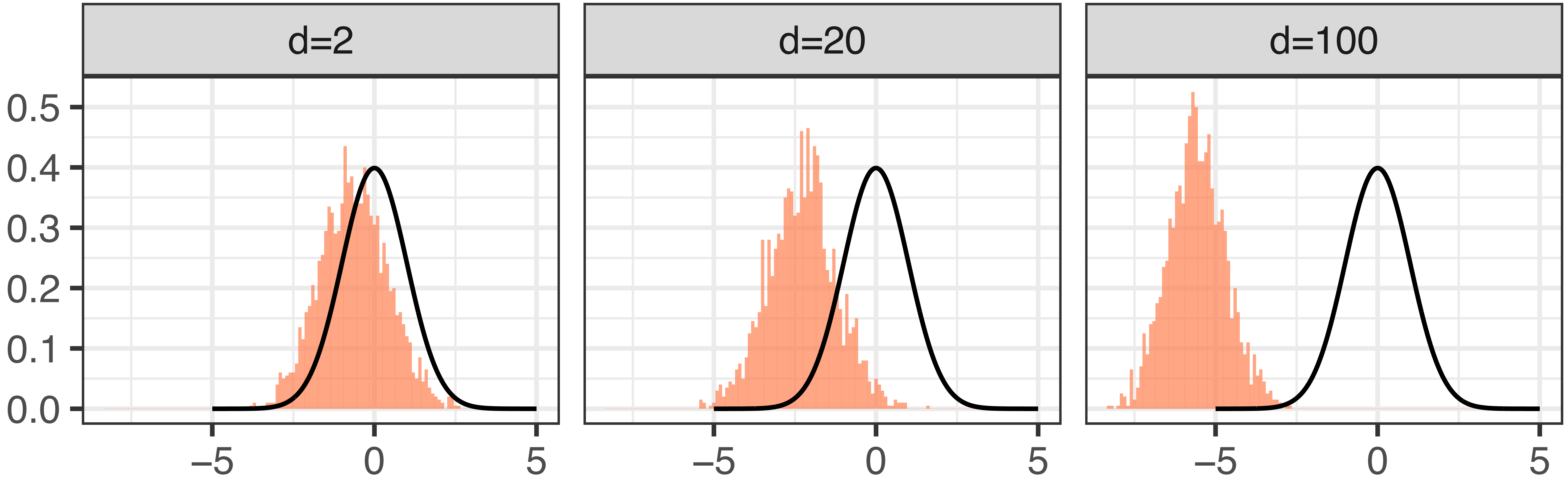
Conservativeness

T. (2025)

We have built a confidence set $\mathrm{CI}_{n,\alpha}$ such that

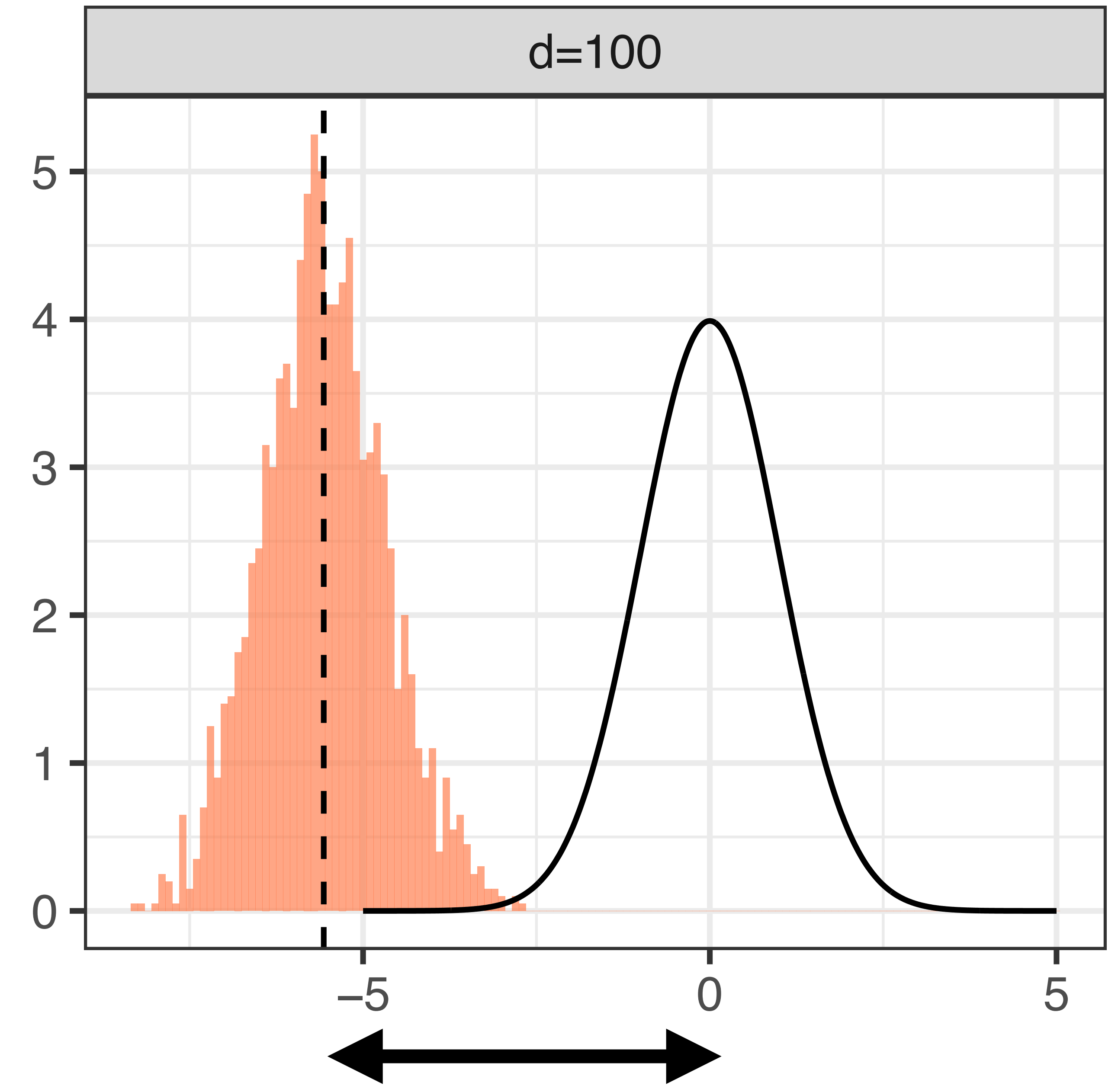
(1) remains valid without the knowledge of the regularity;

(2) shrinks adaptively at a rate depending on the regularity.

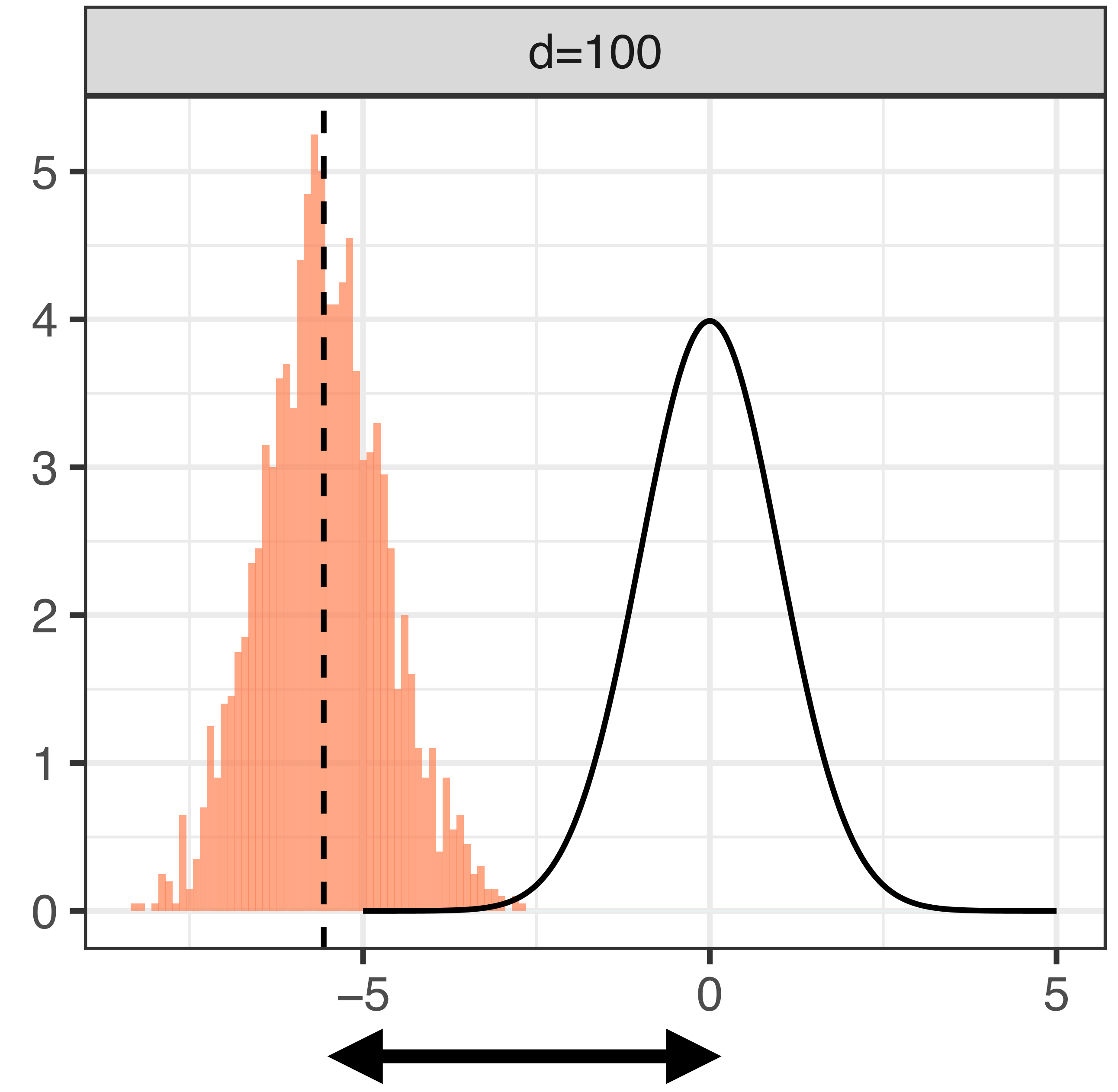


Distribution of $n^{1/2}\bar{\xi}_P / \widehat{\sigma}_P$ for high-dimensional linear regression ($n = 500$)

The undressed bias B_P is the driver of the conservativeness.

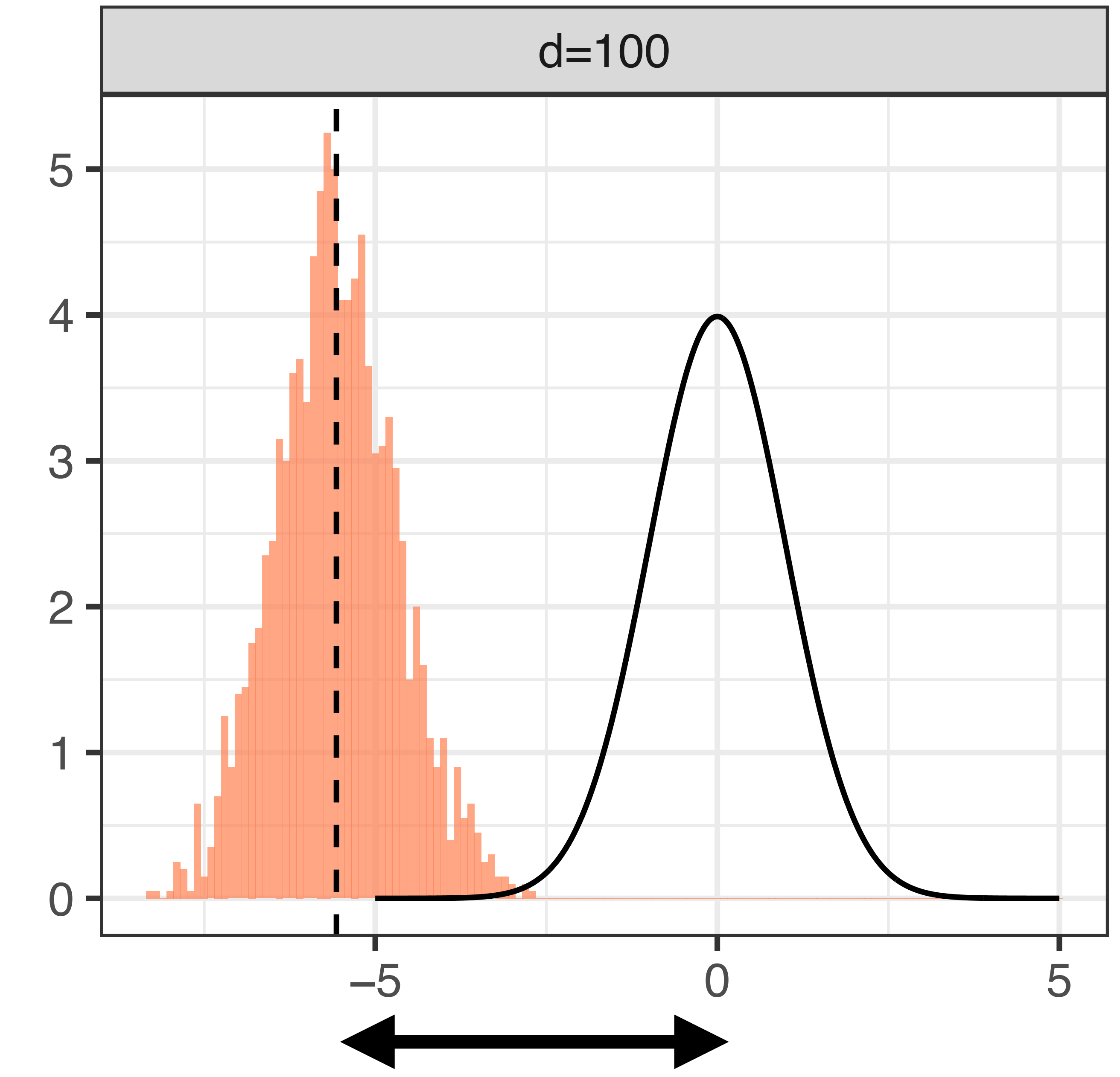


The undressed bias B_P is the driver of the conservativeness.



The undressed bias B_P is the driver of the conservativeness.

With additional assumptions, we *may* be able to construct the estimator $\hat{B}$.

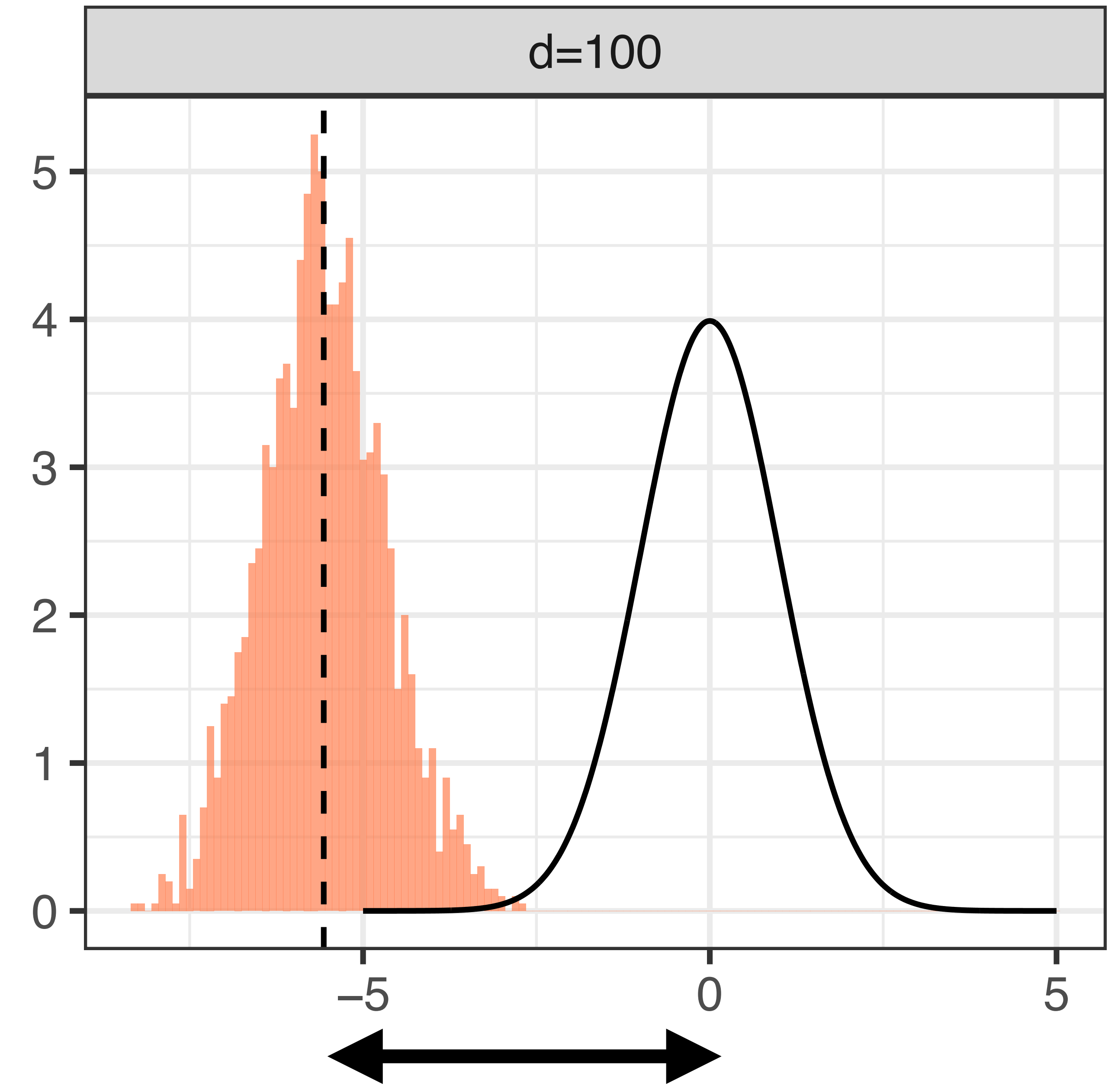


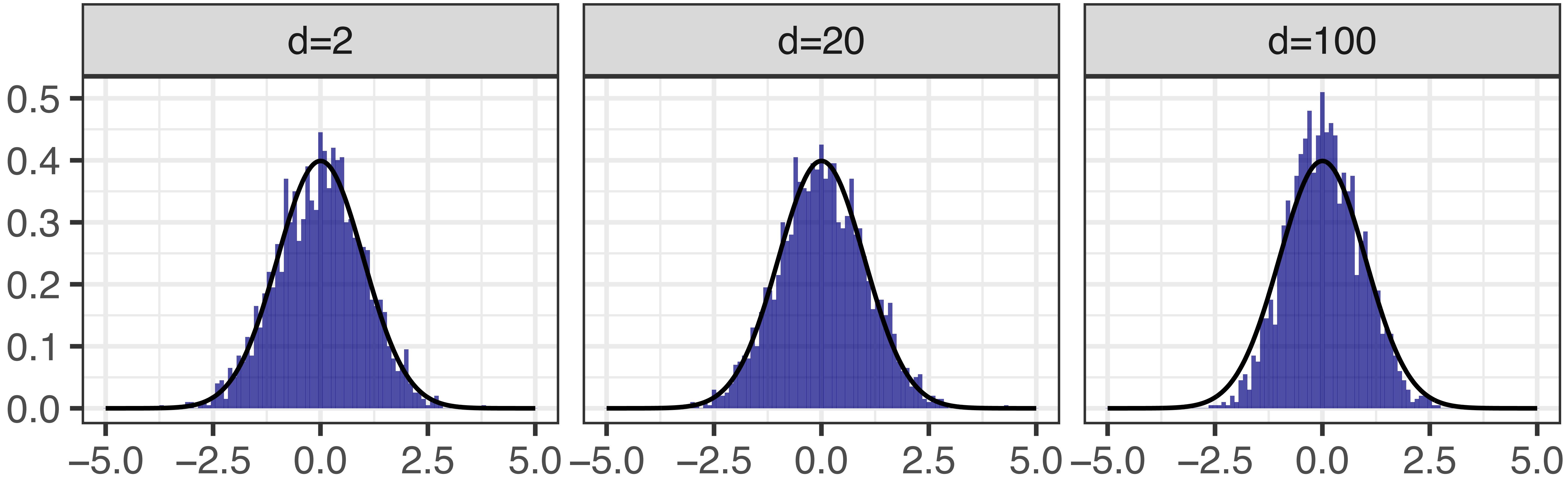
The undressed bias B_P is the driver of the conservativeness.

With additional assumptions, we *may* be able to construct the estimator $\hat{B}$.

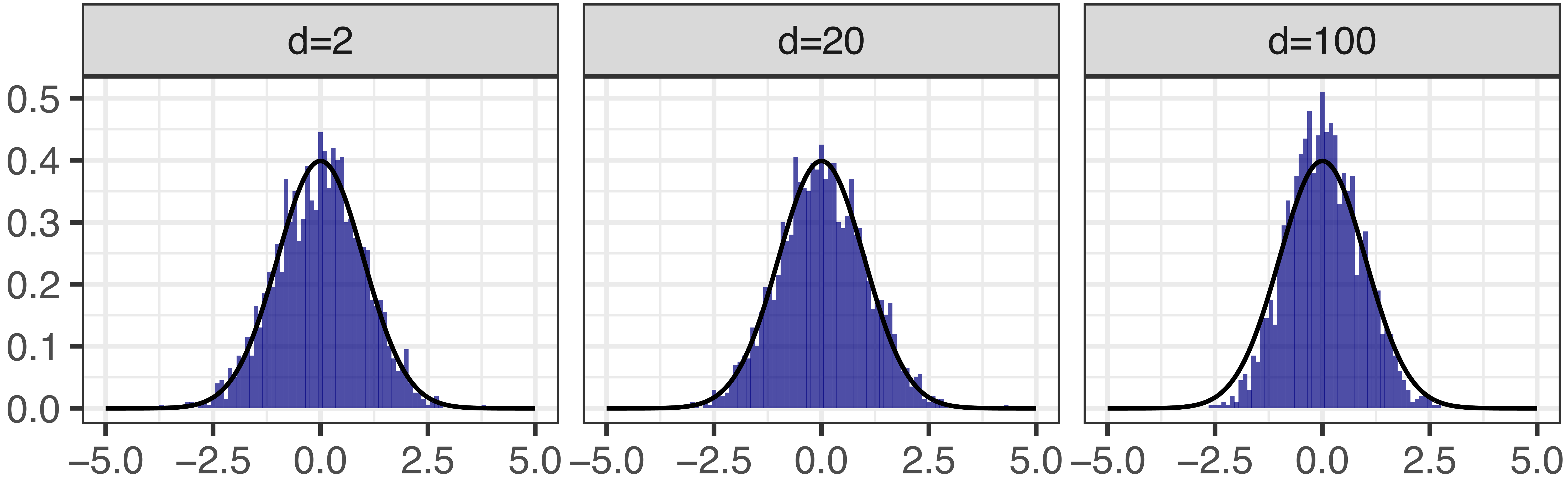
The bias-corrected confidence set is:

$$\mathrm{CI}_{n,\alpha}^{\mathrm{BC}} := \{\theta \in \Theta \;:\; \bar{\xi} + \hat{B} \leq n^{-1/2} z_\alpha \widehat{\sigma}\}.$$





Distribution of $n^{1/2}(\bar{\xi}_P + \hat{B}) / \widehat{\sigma}_P$ for high-dimensional linear regression ($n = 500$)

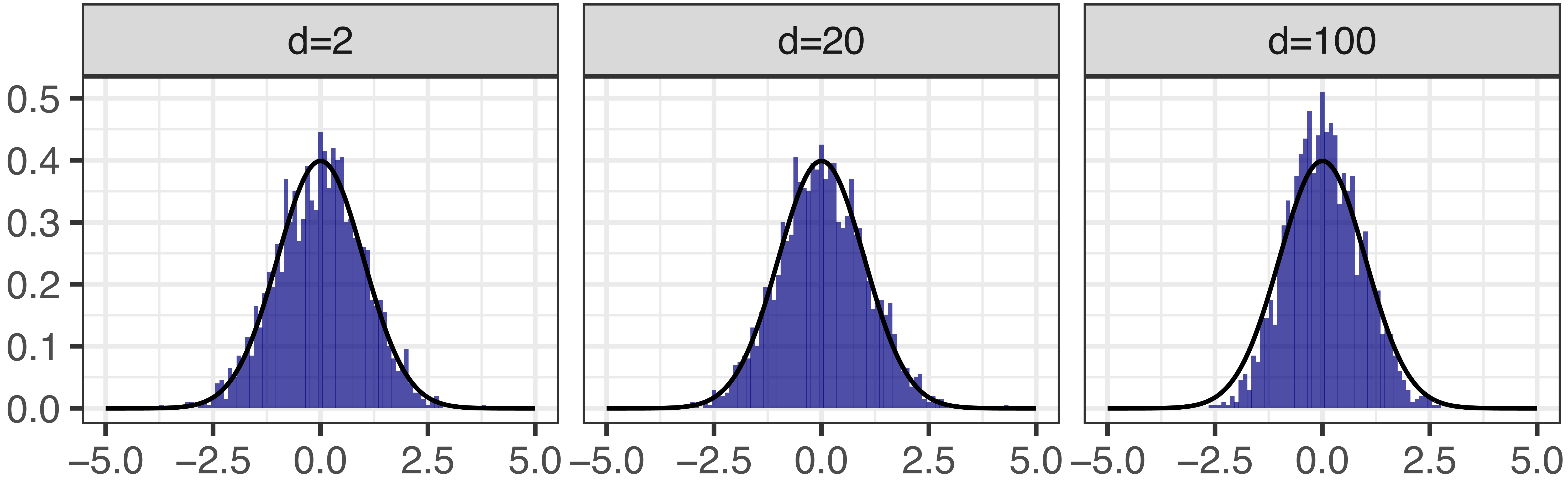


Distribution of $n^{1/2}(\bar{\xi}_P + \hat{B})/\hat{\sigma}_P$ for high-dimensional linear regression ($n = 500$)

Informal

Assuming $\|\hat{\theta}_1 - \theta_P\| \times |\hat{B} - B_P| = o_P(n^{-1/2})$, and additional conditions,

$$\limsup_{n\to\infty} |\mathbb{P}_P(\theta_P \in \mathrm{CI}^{\mathrm{BC}}_{n,\alpha}) - (1-\alpha)| = 0.$$



Distribution of $n^{1/2}(\bar{\xi}_P + \hat{B})/\hat{\sigma}_P$ for high-dimensional linear regression ($n = 500$)

Informal

Assuming $\|\hat{\theta}_1 - \theta_P\| \times |\hat{B} - B_P| = o_P(n^{-1/2})$, and additional conditions,

$$\limsup_{n\to\infty} |\mathbb{P}_P(\theta_P \in \mathrm{CI}^{\mathrm{BC}}_{n,\alpha}) - (1-\alpha)| = 0.$$

The property similar to double robustness emerges.

Thank You

Takatsu, K. and Kuchibhotla, A. K. (2025). Bridging Root-n and Non-standard Asymptotics: Dimension-agnostic Adaptive Inference in M-Estimation, arXiv:2501.07772.

Takatsu, K. (2025). On the Precise Asymptotics of Universal Inference, arXiv:2501.07772.