Adaptive Inference in Irregular M-estimation

Validity, Optimality, and Conservativeness

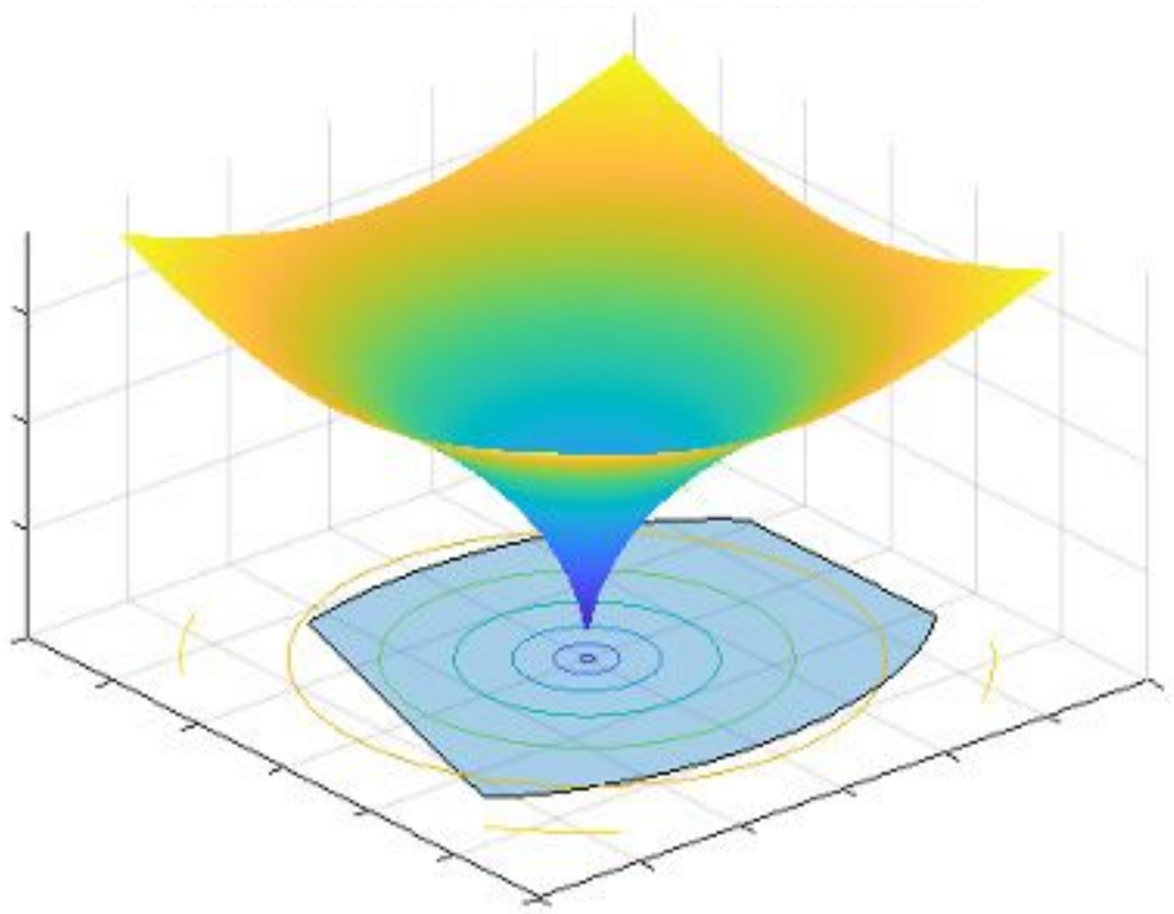
Kenta Takatsu
Based on joint works with Arun Kumar Kuchibhotla
StatDS PhD Research Showcase, April 2025

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

We consider the summary as a minimizer of expected loss fn:

$$P \mapsto \theta_P := \operatorname*{argmin}_{\theta \in \Theta} \mathbb{E}_P[m(X;\theta)].$$

This is called M-estimation



Convex optimization and convex constraints (www.mathworks.com)

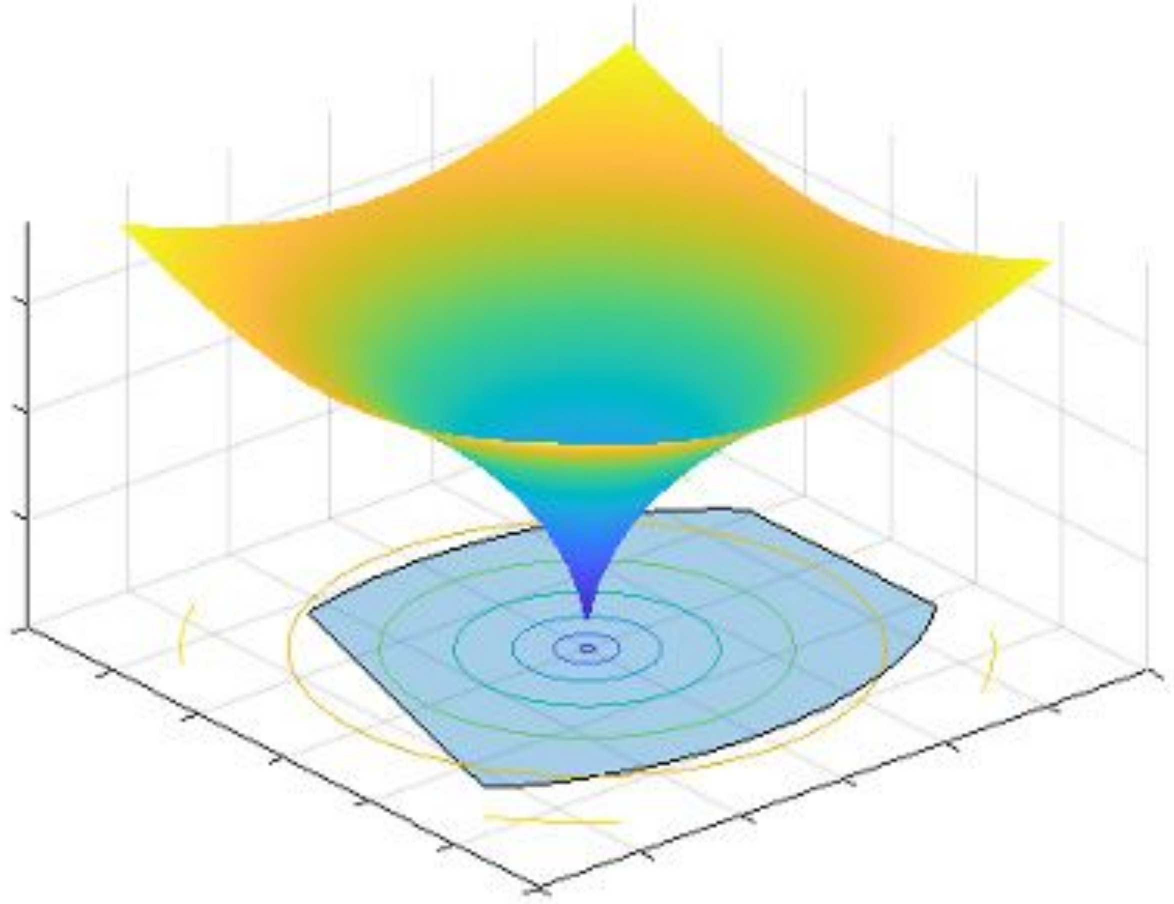
Given observations $\{X_i\}_{i=1}^n$ from a unknown distribution $P \in \mathscr{P}$, we are interested in some "summary" of P.

We consider the summary as a minimizer of expected loss fn:

$$P \mapsto \theta_P := \underset{\theta \in \Theta}{\operatorname{argmin}} \, \mathbb{E}_P[m(X;\theta)].$$

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

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



Convex optimization and convex constraints (www.mathworks.com)

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

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

A "traditional" approach

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

$$\inf_{P \in \mathcal{P}} \mathbb{P}(\theta_P \in \mathrm{CI}_{n,\alpha}) \geq 1 - \alpha.$$

A "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 the 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

$$\inf_{P \in \mathcal{P}} \mathbb{P}(\theta_P \in \mathrm{CI}_{n,\alpha}) \geq 1 - \alpha.$$

A "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 the 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 the Wald Interval

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

[Scheffé and Tukey, 1945; Smirnov, 1952]

Failure of the Wald Interval

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 the Wald Interval

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

Example

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

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

[Scheffé and Tukey, 1945; Smirnov, 1952]

Failure of the Wald Interval

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

Example

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

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

↔

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

[Scheffé and Tukey, 1945; Smirnov, 1952]

Failure of the Wald Interval

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

For many M-estimation defined by

$$\theta_P := \operatorname*{argmin}_{\theta \in \Theta} \mathbb{E}_P[m(X; \theta)],$$

we observe similar irregular behaviors, 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 generally know whether/how the problem is regular or not.

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

Subsampling/Bootstrap typically fail for these problems.

We don't generally know whether/how the problem is regular or not.

Regardless, we show there is 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.

This is adaptive inference

Proposed Procedure

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

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)

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

On the second half, we perform the following:

For each $\theta \in \Theta$:

1. Compute the difference of losses: $\xi_i \equiv \xi_{i,\theta,\widehat{\theta}} := m(X_i;\theta) - m(X_i;\widehat{\theta})$.

2. Include θ in the confidence set if

This is called non-central t-statistics

$$\frac{n^{1/2}\bar{\xi}}{\widehat{\sigma}} \le z_\alpha$$ where $\bar{\xi}$ and $\widehat{\sigma}^2$ are sample mean and variance of $\{\xi_i\}_{i=1}^n$.

Proposed Procedure

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

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

On the second half, we perform the following:

For each $\theta \in \Theta$:

1. Compute the difference of losses: $\xi_i \equiv \xi_{i,\theta,\widehat{\theta}} := m(X_i;\theta) - m(X_i;\widehat{\theta})$.

2. Include θ in the confidence set if

   $\dfrac{n^{1/2}\bar{\xi}}{\widehat{\sigma}} \le z_\alpha$ where $\bar{\xi}$ and $\widehat{\sigma}^2$ are sample mean and variance of $\{\xi_i\}_{i=1}^n$.

The final confidence set is $\mathrm{CI}_{n,\alpha} := \left\{\theta \in \Theta \,:\, n^{1/2}\widehat{\sigma}^{-1}\bar{\xi} \le z_\alpha\right\}$.

Why does this work?

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* $\widehat{\theta} \in \Theta$.

Why does this work?

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 $\left\{\theta \in \Theta \;:\; \mathbb{E}_P[m(X;\theta) - m(X;\hat{\theta}) \mid \hat{\theta}] \leq 0\right\}$.

Why does this work?

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.

Why does this work?

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 $\gamma_{n,\alpha}$.

A Brief History

Inverting the risk of an irregular estimator is not a new idea.

A Brief History

Inverting the risk of an irregular estimator is not a new idea.

Stein mentioned the idea in passing.

1981

[Stein, 1981]

A Brief History

Inverting the risk of an irregular estimator is not a new idea.

The inversion based on CLT appeared in nonparametrics in the late 1990s.

1981

1996~1998

[Beran, 1996; Beran and Dümbgen, 1998]

A Brief History

Inverting the risk of an irregular estimator is not a new idea.

Robins and van der Vaart combine the CLT and sample-splitting.

1981

1996~1998

2006

[Robins and van der Vaart, 2006]

A Brief History

Inverting the risk of an irregular estimator is not a new idea.

Many works from CMU use this idea for irregular inference.

1981

1996~1998

2006

2019 ~

[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 is not a new idea.

Literature in stochastic programming uses CLT for general loss + constraints.

“Universal confidence set” (Vogel, 2008) but without sample-splitting.

1981

1989~1996

1996~1998

2006

2008

2019 ~

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

A Brief History

Inverting the risk of an irregular estimator is not a new idea.

[Wasserman et al. (2020)]

Properties of the Confidence Set

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

Reminder: $\mathrm{CI}_{n,\alpha} := \left\{\theta \in \Theta \,:\, n^{1/2}\widehat{\sigma}^{-1}\overline{\xi} \le z_\alpha\right\}.$

Validity

Size of the CI

Properties of the Confidence Set

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

Reminder: $\mathrm{CI}_{n,\alpha} := \left\{\theta \in \Theta \,:\, n^{1/2}\widehat{\sigma}^{-1}\overline{\xi} \leq z_\alpha\right\}.$

Validity

Size of the CI

Validity holds when θ_P is not unique.

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

Relatively mild regularity is required for the CLT.

Properties of the Confidence Set

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

Reminder: $\mathrm{CI}_{n,\alpha} := \left\{\theta \in \Theta \,:\, n^{1/2}\widehat{\sigma}^{-1}\overline{\xi} \le z_\alpha\right\}.$

Validity

Validity holds when θ_P is not unique.

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

Relatively mild regularity is required for the CLT.

Size of the CI

The CI shrinks to a singleton only when θ_P is unique.

The diameter shrinks 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)

Reminder: $\mathrm{CI}_{n,\alpha} := \left\{\theta \in \Theta \,:\, n^{1/2}\widehat{\sigma}^{-1}\overline{\xi} \leq z_\alpha\right\}.$

Validity

Validity holds when θ_P is not unique.

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

Relatively mild regularity is required for the CLT.

Size of the CI

The CI shrinks to a singleton only when θ_P is unique.

The diameter shrinks 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 problems; Manski's maximum score estimator; Quantile; Argmin.

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 $\eta < 1 + \beta$.

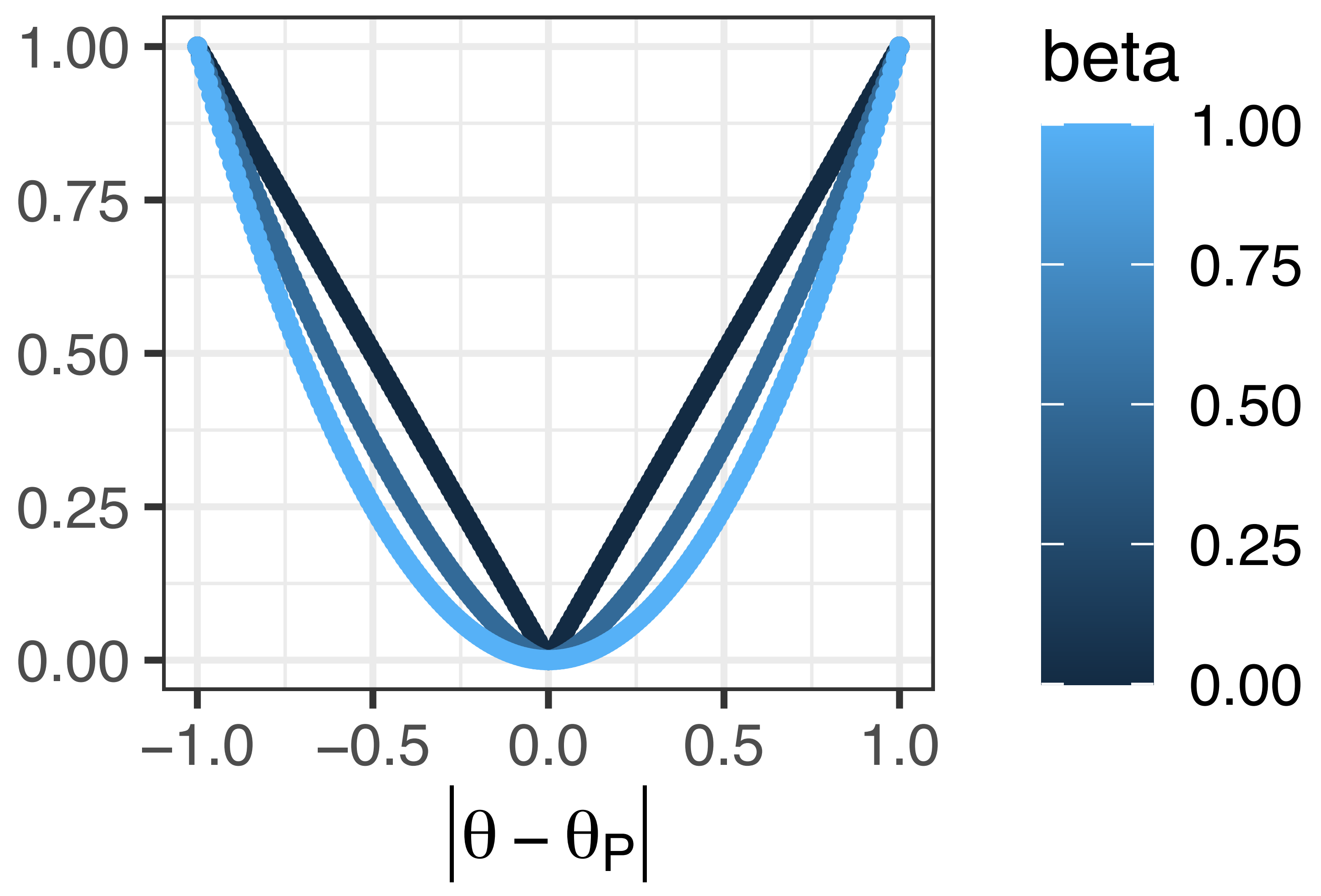


Illustration of curvature

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 8 (informal)

The diameter of the confidence set satisfies

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

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

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 8 (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 shrinks adaptively to unknown β and η.

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

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 8 (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, the moments of the local envelope $\sup_{\|\theta-\theta_P\|<\delta} |m(X;\theta) - m(X;\theta_P)|$.

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

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 8 (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 quality of the initial estimator

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

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

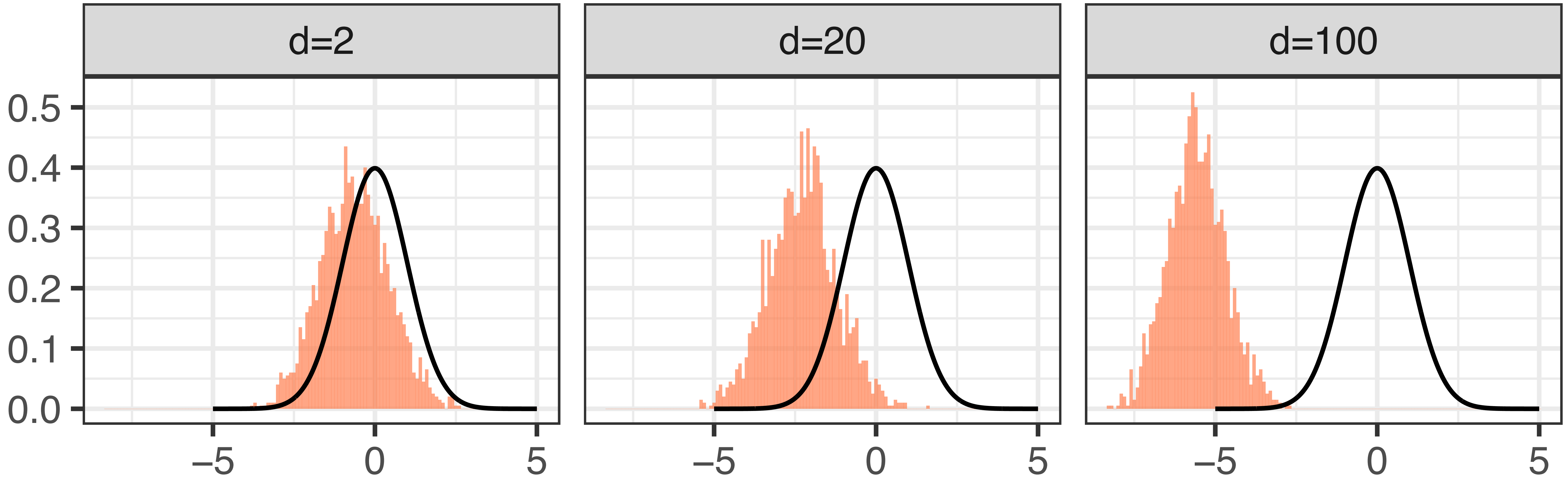
Conservativeness

T. (2025)

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

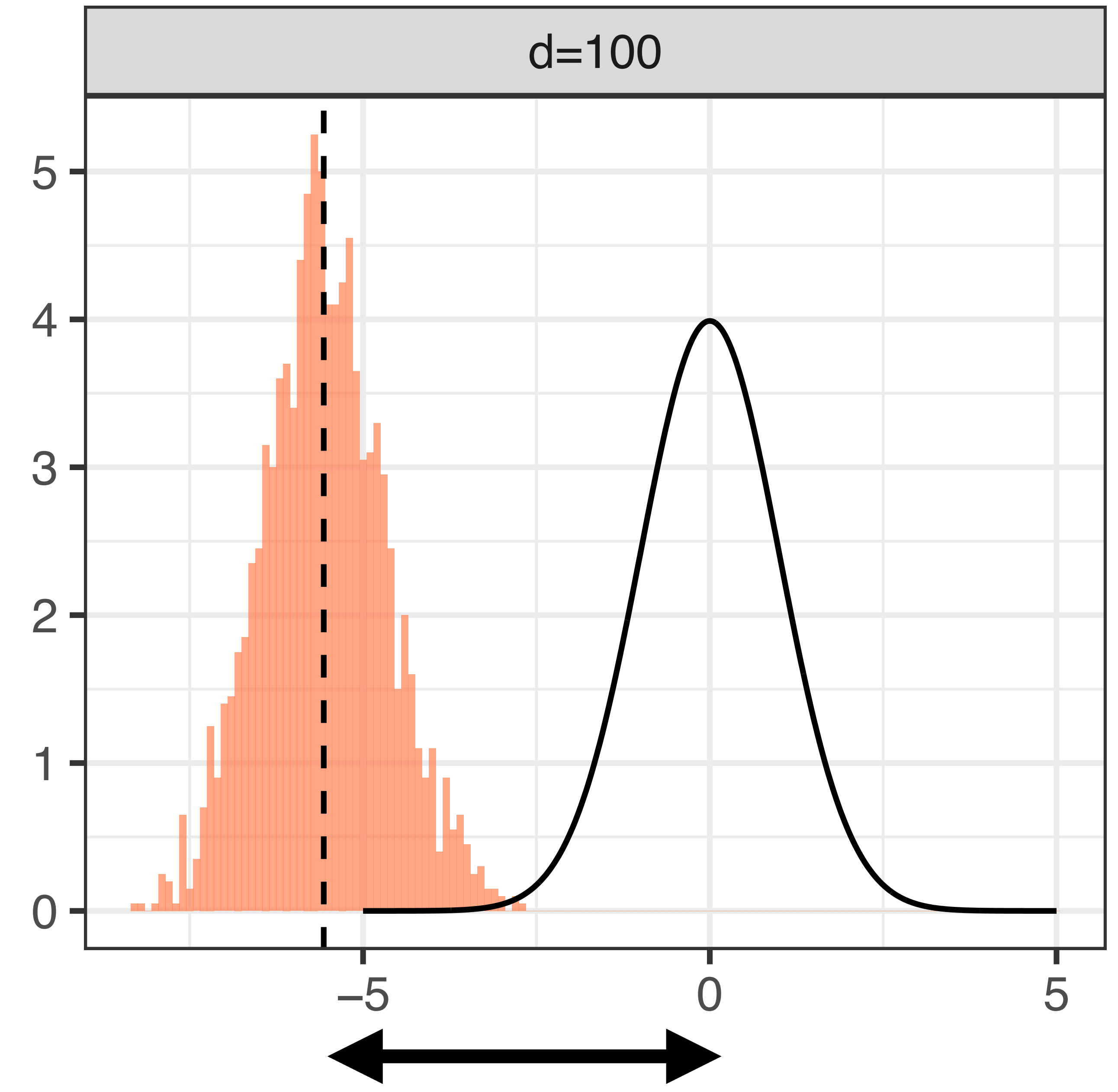
(1) remains valid without the knowledge of the regularity;

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

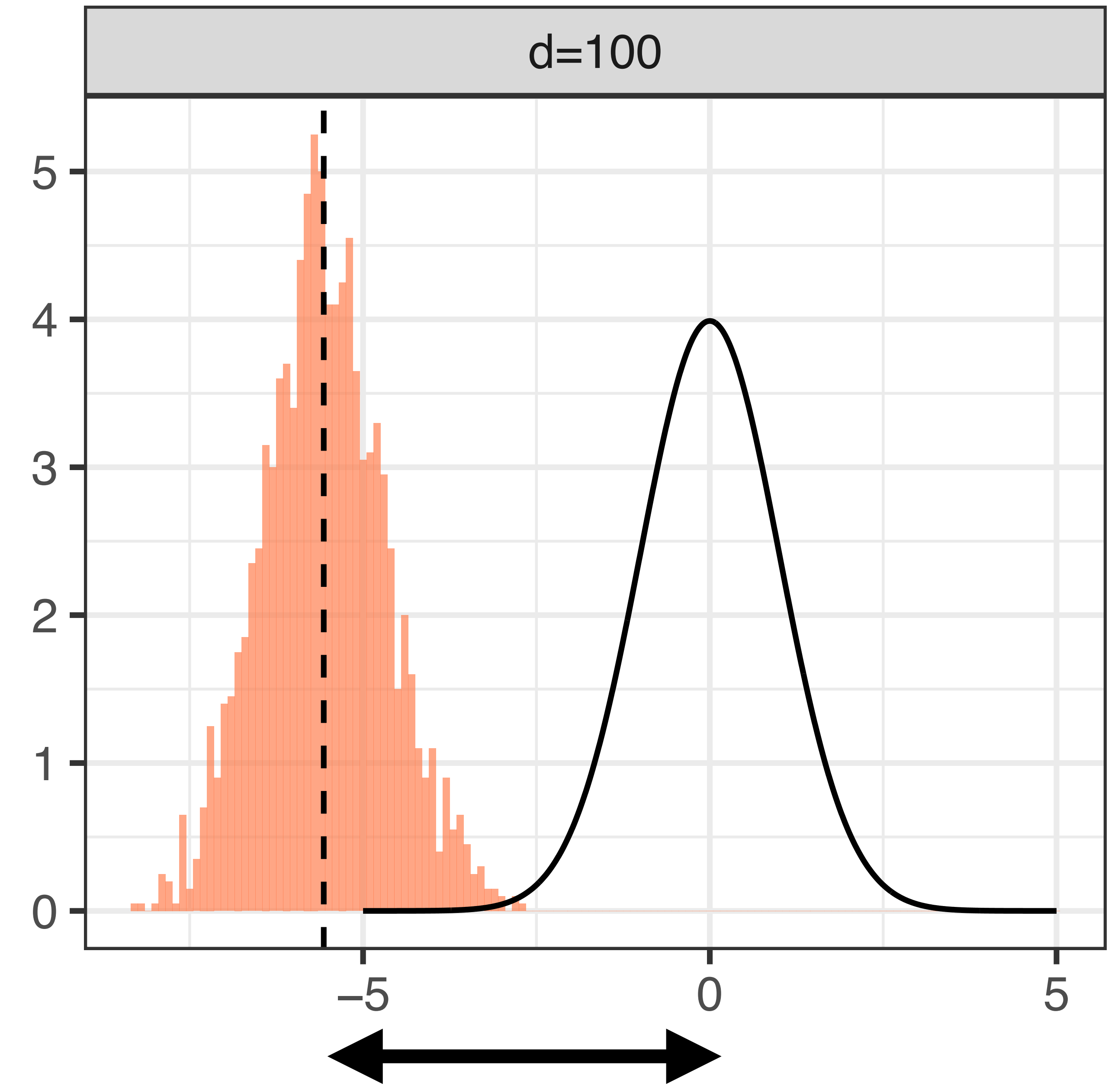


Distribution of $n^{1/2}\bar{\xi}_P / \widehat{\sigma}_P$ for high-dimensional linear regression $(n = 500)$.

The unaddressed bias B_P is the driver of conservativeness.

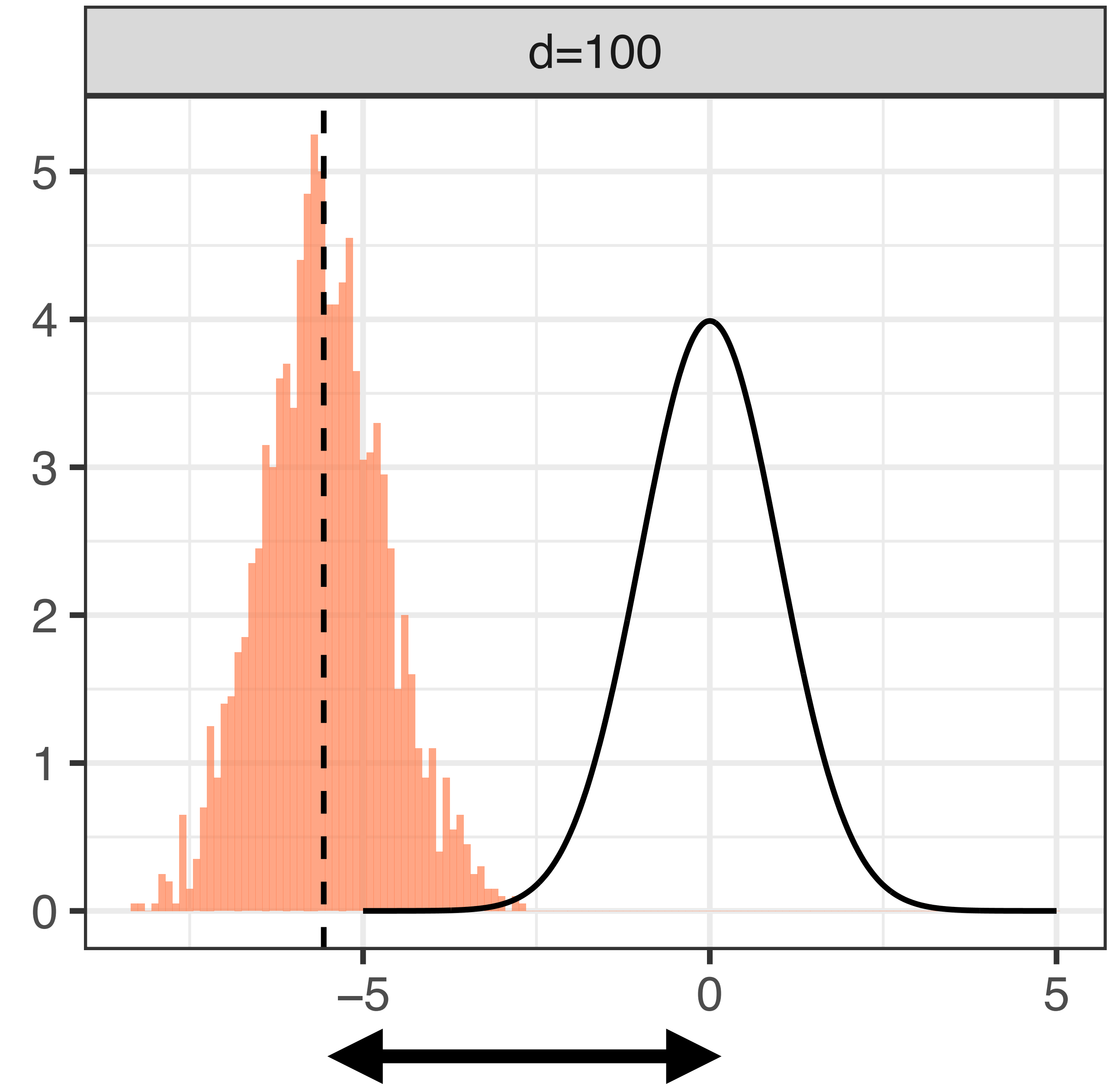


The unaddressed bias B_P is the driver of conservativeness.



The unaddressed bias B_P is the driver of conservativeness.

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

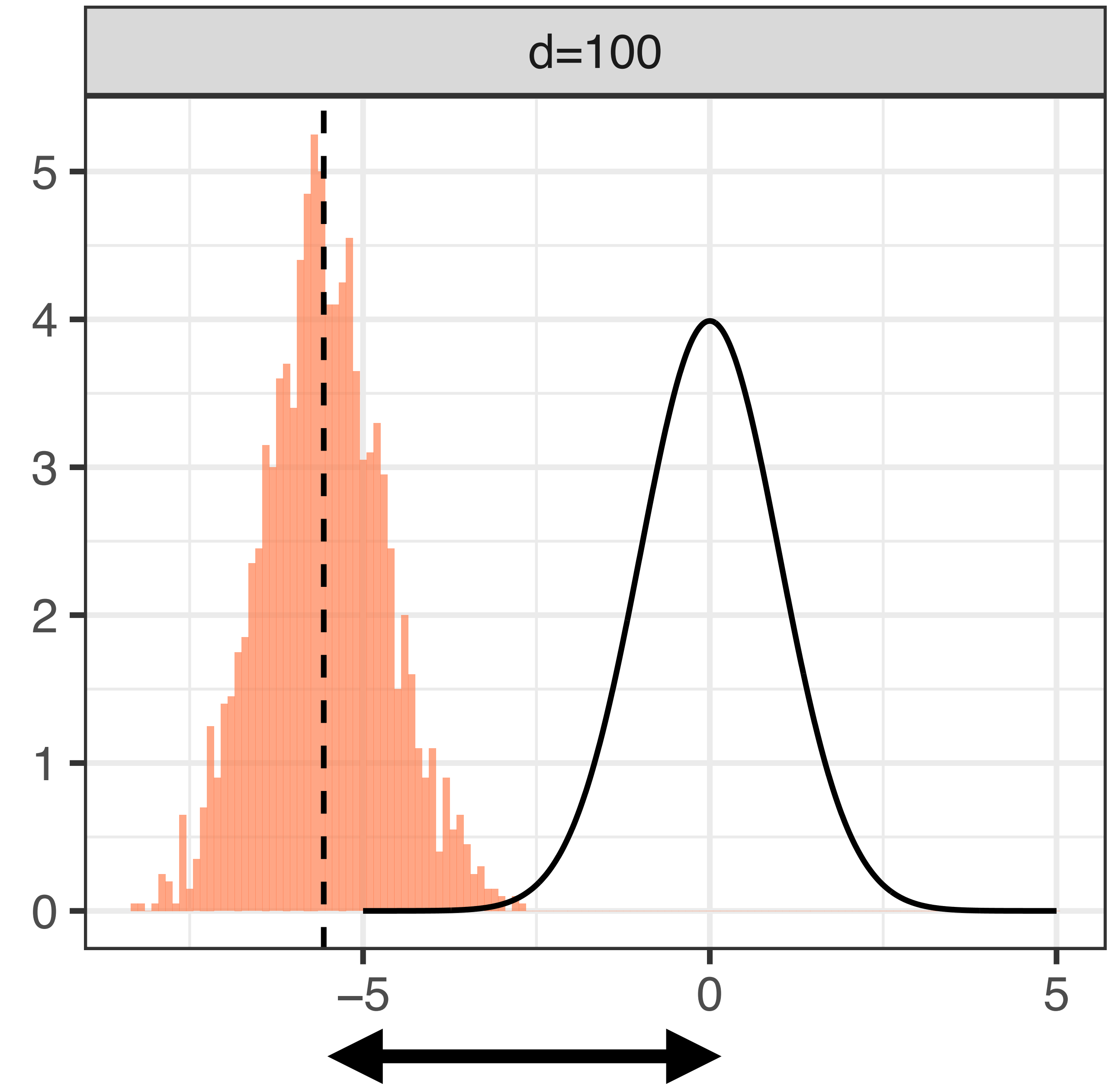


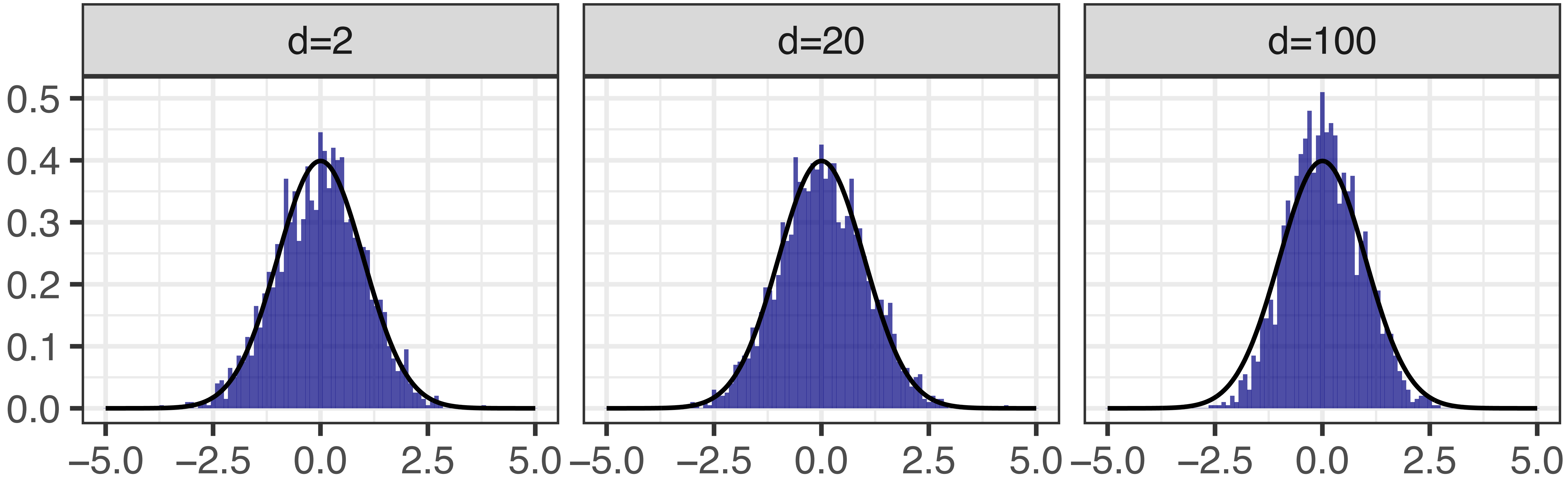
The unaddressed bias B_P is the driver of conservativeness.

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

The bias-corrected confidence set is:

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



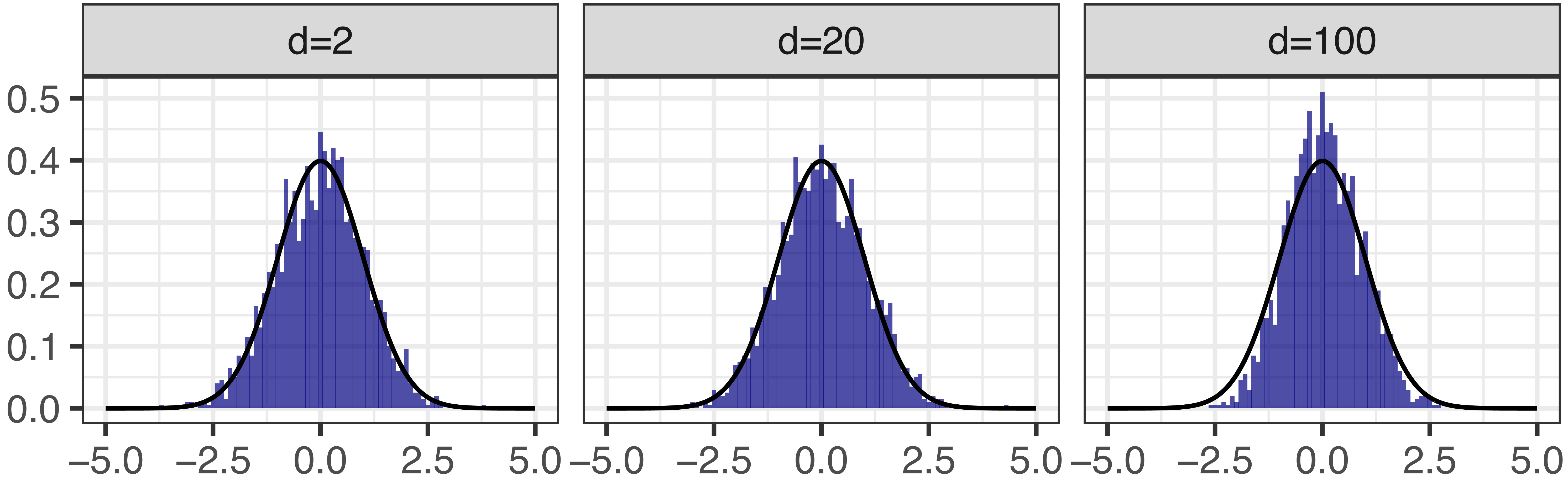


Distribution of $n^{1/2}(\bar{\xi}_P + \hat{B})/\hat{\sigma}_P$ for high-dimensional linear regression ($n = 500$).

Theorem (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 ($n = 500$).

Theorem (Informal)

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

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

A property similar to double robustness emerges.

Summary

Combining the CLT for t-statistics and sample-splitting provides a general confidence set for M-estimation.

The confidence set is valid under very weak assumptions.

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

Avoiding conservativeness requires additional efforts, such as bias-correction. For some problems, the requirement looks similar to double robustness from semiparametric theory.

Open Problem

Can we use this framework for the profile likelihood:

$$P \mapsto \theta_P := \operatorname*{argmin}_{\theta \in \Theta} \min_{\eta \in \mathscr{H}} \mathbb{E}_P[m(X; \theta, \eta)]$$

where $\Theta \in \mathbb{R}^d$ and $\mathscr{H}$ is an inner product space?

Proportional hazard model

Partial linear regression

Single index model

Casual functional

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:2503.14717.

Uniform Validity

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

Q. What is required for the validity of $\mathrm{CI}_{n,\alpha} := \{\theta \in \Theta \,:\, \bar{\xi} \leq n^{-1/2} z_\alpha \widehat{\sigma}\}$?

Uniform Validity

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

Q. What is required for the validity of $\mathrm{CI}_{n,\alpha} := \{\theta \in \Theta \,:\, \bar{\xi} \le n^{-1/2} z_\alpha \hat{\sigma}\}$?

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 Kolmogorov-Smirnov distance by

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

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

Uniform Validity

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

Q. What is required for the validity of $\mathrm{CI}_{n,\alpha} := \{\theta \in \Theta : \bar{\xi} \leq n^{-1/2} z_\alpha \hat{\sigma}\}$?

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 Kolmogorov-Smirnov distance by

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

This measures *distance* between the the t-statistics and standard Normal

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

Uniform Validity

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

Q. What is required for the validity of $\mathrm{CI}_{n,\alpha} := \{\theta \in \Theta \ : \ \bar{\xi} \le n^{-1/2} z_\alpha \hat{\sigma}\}$?

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 Kolmogorov-Smirnov distance by

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

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

Theorem

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