Rate-adaptive robustly valid confidence sets for M-/Z-estimation problems

Arun Kumar Kuchibhotla

16 June, 2025

Carnegie Mellon University

Collaborators

Kenta Takatsu (CMU)

Woonyoung Chang (CMU)

Table of contents

[1] Joint work with Kenta Takatsu (arXiv:2501.07772)
[2] Joint work with Woonyoung Chang (arXiv:2407.12278)

Traditional inference framework

Inference: confidence intervals

- ⋆ The construction of confidence sets for functionals is a standard problem in statistics.
- ⋆ Suppose $\theta(P), P \in \mathcal{P}$ is a functional of interest, for example, the mean of P or a coefficient in a regression model.
- ⋆ Traditional inference methods such as Wald or resampling (e.g. bootstrap or subsampling) proceed as follows.
- ⋆ Assuming the existence of an estimator $\widehat{\theta}_n$ based on n observations such that

$$r_n(\widehat{\theta}_n - \theta(P)) \xrightarrow{d} L,$$

a confidence interval can be constructed as

$$\widehat{\mathrm{CI}}_{n,\alpha} := \left[\widehat{\theta}_n - \frac{\widehat{q}_{1-\alpha/2}}{\widehat{r}_n}, \widehat{\theta}_n - \frac{\widehat{q}_{\alpha/2}}{\widehat{r}_n}\right],$$

where $\widehat{q}_\gamma$ represents an estimate of the γ-th quantile of the random variable L, and $\widehat{r}_n$ is an estimate of r_n, if unknown.

Example: Linear Regression (fixed d)

- ⋆ Suppose $(X, Y) \in \mathbb{R}^{d+1}$ is a random vector from a distribution P and we are interested in the projection parameter $\theta_0 = \theta(P)$ defined

$$\theta(P) = \underset{\theta \in \mathbb{R}^d}{\arg\min}\ \mathbb{E}_P[(Y - X^\top\theta)^2].$$

- ⋆ Because of unconstrained optimization, $\theta(P)$ is also the solution to the equation

$$\mathbb{E}_P[X(Y - X^\top\theta)] = 0.$$

- ⋆ Using IID data $(X_i, Y_i), 1 \le i \le n$, $\theta(P)$ can be estimated using

$$\widehat{\theta}_n = \underset{\theta \in \mathbb{R}^d}{\arg\min} \sum_{i=1}^{n} (Y_i - X_i^\top\theta)^2.$$

- ⋆ For a fixed d, assuming the invertibility of $\Sigma = \mathbb{E}[XX^\top]$, as $n \to \infty$,

$$n^{1/2}(\widehat{\theta}_n - \theta(P)) \overset{d}{\to} N(0, \Sigma^{-1}V\Sigma^{-1}),$$

where $V = \mathbb{E}[XX^\top(Y - X^\top\theta(P))^2]$; no linear model or Gaussianity.

Wald Inference: Linear Regression (fixed d)

- ⋆ The asymptotic variance can be estimated as $\widehat{\Sigma}^{-1}\widehat{V}\widehat{\Sigma}^{-1}$.
- ⋆ For any vector $c \in \mathbb{R}^d$, the Wald confidence interval for $c^\top\theta(P)$ can be obtained as

$$\widetilde{\mathrm{CI}}_{n,\alpha}^{\mathrm{Wald}}(c) \;:=\; \left[c^\top\widehat{\theta}_n \pm z_{\alpha/2}\left(\frac{c^\top\widehat{\Sigma}^{-1}\widehat{V}\widehat{\Sigma}^{-1}c}{n}\right)^{1/2}\right].$$

- ⋆ Again with d fixed, as $n \to \infty$, this confidence interval has an asymptotic coverage of $1-\alpha$.
- ⋆ This nicety fails when dimensions grow rapidly or when constraints are placed on the projection parameter.

Failure of traditional inference: Increasing dimension

Asymptotics: Increasing dimension

- ⋆ With some algebraic manipulation, the OLS estimator satisfies

$$\widehat{\theta}_n - \theta(P) = \frac{1}{n}\sum_{i=1}^{n} \widehat{\Sigma}^{-1} X_i (Y_i - X_i^\top \theta(P)).$$

- ⋆ Asymptotic normality is claimed, for fixed d, by replacing $\widehat{\Sigma}^{-1}$ with Σ^{-1} with "negligible" error:

$$\begin{aligned}\widehat{\theta}_n - \theta(P) = &\frac{1}{n}\sum_{i=1}^{n} \Sigma^{-1} X_i (Y_i - X_i^\top \theta(P)) \\ &+ \frac{1}{n}\sum_{i=1}^{n} (\widehat{\Sigma}^{-1} - \Sigma^{-1}) X_i (Y_i - X_i^\top \theta(P)).\end{aligned}$$

- ⋆ The first term is mean zero and responsible for asymptotic normality, and for fixed d, the second term is negligible compared to the first.
- ⋆ Keeping track of dimension dependence, the first term is of order $1/\sqrt{n}$ and the second is of order d/n, along any direction.

Asymptotics: Increasing dimension

- ⋆ If $d = \tilde{o}(n^{1/2})$, then

$$n^{1/2}(\widehat{\theta}_n - \theta(P)) \overset{d}{\approx} N(0, \Sigma^{-1} V \Sigma^{-1}).$$

The asymptotic variance can be consistently estimated as if d were fixed. (Here $\tilde{o}$ hides poly $\log n$ factors.)

- ⋆ If $d = \tilde{o}(n^{2/3})$, then

$$n^{1/2}(\widehat{\theta}_n - \theta(P) - \mathcal{B}(P)) \overset{d}{\approx} N(0, \Sigma^{-1} V \Sigma^{-1}),$$

where

$$\mathcal{B}(P) = -\frac{1}{n^2} \sum_{i=1}^{n} \Sigma^{-1} X_i (Y_i - X_i^\top \theta(P)) \|X_i\|_{\Sigma^{-1}}^2.$$

- ⋆ Unsurprisingly, $\mathcal{B}(P) = o_p(n^{-1/2})$ if $\mathbb{E}[Y|X] = X^\top \theta(P)$ and $d = o(n)$.

Inference with increasing dimension

- ⋆ This implies that, even if the bias is known, we have asymptotic normality only when $d \ll n^{2/3}$.

- ⋆ Chang, Kuchibhotla, and Rinaldo (2023, ArXiv:2307.00795) proposed a plug-in bias estimator $\widehat{B}_n$ so that

$$n^{1/2}(\widehat{\theta}_n - \widehat{B}_n - \theta(P)) \overset{d}{\approx} N(0, \Sigma^{-1} V \Sigma^{-1}), \quad \text{whenever} \quad d = \tilde{o}(n^{2/3}),$$

  along any direction. They also proved the consistency of the classical variance estimator.

- ⋆ Lin et al. (2024, ArXiv:2411.02909) proposed a jackknife estimator with similar properties for a general Z-estimator under a restrictive compact parameter space assumption.

- ⋆ Hence, traditional Wald inference is only valid for $d = \tilde{o}(n^{2/3})$. We do not know of an asymptotically normal estimator when $d \gg n^{2/3}$.

Failure of traditional inference: Constraints

With constraints

- ⋆ The situation is much worse with constraints, even if d is fixed as $n \to \infty$.
- ⋆ Suppose

$$\theta(P) = \underset{\theta \in \Theta}{\arg\min}\, \mathbb{E}[(Y - X^\top \theta)^2],$$

  for some set $\Theta \subseteq \mathbb{R}^d$.
- ⋆ The limiting distribution of the sample estimator $\widehat{\theta}_n$ is highly dependent on the regularity of $\theta(P)$ with respect to Θ (e.g., equality of different notions of tangent cones).
- ⋆ The limit could be a projected Gaussian; see Pflug (1995), Geyer (1994), and Shapiro (2000).

Examples with constraints

- ⋆ Examples with constraints are relevant in practice.

- ⋆ **Sparsity** inducing least squares:

$$\Theta = \{\theta \in \mathbb{R}^d : \|\theta\|_0 \leq k\},$$

or

$$\Theta = \{\theta \in \mathbb{R}^d : \|\theta\|_1 \leq t\},$$

or

$$\Theta = \left\{\theta \in \mathbb{R}^d : \sum_{j=1}^{k} \|\theta_{G_j}\|_2 \leq t\right\}.$$

- ⋆ **Shape** inducing least squares:

$$\Theta = \{\theta \in \mathbb{R}^d : \theta \succeq 0\},$$

or

$$\Theta = \{\theta \in \mathbb{R}^d : \Delta_1\theta \succeq 0\},$$

where $\Delta_1\theta$ yields the first order differences of θ; e.g., $(\Delta_1\theta)_1 = \theta_2 - \theta_1$.

Approach 1: Risk inversion for M-estimation[a]

[a] Joint work with Kenta Takatsu (arXiv:2501.07772)

The Setting

- $\star$ For any distribution $P \in \mathcal{P}$, define the functional of interest

$$\theta(P) \;:=\; \arg\min_{\theta \in \Theta} \mathbb{M}_P(\theta), \quad \mathbb{M}_P(\theta) := \mathbb{E}_P[m(Z;\theta)].$$

  Here Θ can be infinite-dimensional and $\theta(P)$ can be a set.

- $\star$ Clearly, for any estimator $\widetilde{\theta} \in \Theta$,

$$\theta(P) \subseteq \left\{\theta \in \Theta :\, \mathbb{M}(\theta) - \mathbb{M}(\widetilde{\theta}) \le 0\right\}.$$

- $\star$ Of course, the right hand set is not computable based on the data. However, we can construct two sets on the basis of this intuition.

- $\star$ **Weak miscoverage:**

$$\mathrm{MC}(\widehat{\mathrm{CI}}_{n,\alpha}) = \sup_{\theta^* \in \theta(P)} \mathbb{P}_P\left(\theta^* \notin \widehat{\mathrm{CI}}_{n,\alpha}\right).$$

Confidence Sets

- ⋆ Randomly split $\{1, 2, \ldots, 2n\}$ into two equal parts $\mathcal{I}_1$ and $\mathcal{I}_2$.
- ⋆ Compute any estimator $\widetilde{\theta}_1$ using $\{Z_i : i \in \mathcal{I}_1\}$.
- ⋆ On the second half, set $\widehat{\mathbb{M}}_n(\theta) = n^{-1} \sum_{i\in\mathcal{I}_2} m(Z_i; \theta)$ and compute

$$\begin{aligned} \widehat{\mathrm{CI}}_n^{\dagger} &:= \left\{\theta \in \Theta : \widehat{\mathbb{M}}_n(\theta) - \widehat{\mathbb{M}}_n(\widetilde{\theta}_1) \;\le\; 0\right\}, \\ \widehat{\mathrm{CI}}_{n,\alpha} &:= \left\{\theta \in \Theta : \widehat{\mathbb{M}}_n(\theta) - \widehat{\mathbb{M}}_n(\widetilde{\theta}_1) \;\le\; \frac{z_{\alpha/2}}{n^{1/2}}\widehat{\sigma}(\theta)\right\}, \end{aligned} \tag{1}$$

  where $\widehat{\sigma}(\theta)$ is the standard deviation of $m(Z_i; \theta) - m(Z_i; \widetilde{\theta}_1)$, $i \in \mathcal{I}_2$.
- ⋆ Clearly,

$$\widehat{\mathrm{CI}}_n^{\dagger} \subseteq \widehat{\mathrm{CI}}_{n,\alpha} \quad \text{for any} \quad \alpha \in (0, 1), n \ge 1.$$

- ⋆ This is essentially risk inversion and is an old idea.

History

Statistics	Operations Research
(1) Without sample splitting	
Stein (1981)	Pflug (1991, 1995, 2003)
Li (1981)	Vogel (2008)
Beran (1996)	Vogel and Seeger (2017)
Geyer (1996)	Guigues et al. (2017)
Beran and Dumbgen (1998)	
(2) With sample splitting	
Robins and van der Vaart (2006)	
Hoffmann and Nickl (2011)	
Carpentier (2013)	
Chakravarti et al. (2019)	
Kim and Ramdas (2024)	
Park et al. (2023)	

Validity

- ⋆ For any $\widetilde{\theta}_1$, we have

$$\mathrm{MC}(\widehat{\mathrm{CI}}_{n,\alpha}) \;\le\; \mathrm{MC}(\widehat{\mathrm{CI}}_n^{\dagger}) \;\le\; \sup_{\theta^*\in\theta(P)} \mathbb{E}\left[\frac{\sigma^2(\widetilde{\theta}_1)}{\sigma^2(\widetilde{\theta}_1) + n\mathbb{C}^2(\widetilde{\theta}_1)}\right],$$

where

$$\sigma^2(\theta') := \mathrm{Var}_P(m(\theta^*, Z) - m(\theta', Z)),$$
$$\mathbb{C}(\theta') := \mathbb{E}[m(\theta, Z)] - \min_{\theta\in\Theta} \mathbb{E}[m(\theta, Z)].$$

- ⋆ Note that if $\widetilde{\theta}_1$ is inconsistent for $\theta(P)$, i.e., $\mathrm{dist}(\widetilde{\theta}_1, \theta(P)) \neq o_p(1)$ as $n \to \infty$, then

$$\mathrm{MC}(\widehat{\mathrm{CI}}_{n,\alpha}) \;\le\; \mathrm{MC}(\widehat{\mathrm{CI}}_n^{\dagger}) \;\to\; 0, \quad \text{as} \quad n \to \infty.$$

- ⋆ In fact, if $\sigma^2(\theta') \le \|\theta^* - \theta'\|^{2\eta}$ and $\mathbb{C}(\theta') \ge \|\theta^* - \theta'\|^{1+\beta}$, then

$$\mathrm{MC}(\widehat{\mathrm{CI}}_n^{\dagger}) \le \sup_{\theta^*\in\theta(P)} \mathbb{E}\left[\left(1 + n\|\widetilde{\theta}_1 - \theta^*\|^{2+2\beta-2\eta}\right)^{-1}\right].$$

Validity

- ⋆ Hence, $\widehat{\mathrm{CI}}_n^{\dagger}$ is an asymptotically uniformly conservative confidence set if $\widetilde{\theta}_1$ converges at a slow enough rate to $\theta(P)$.

- ⋆ On the other hand, $\widehat{\mathrm{CI}}_{n,\alpha}$ also satisfies

$$\mathrm{MC}(\widehat{\mathrm{CI}}_{n,\alpha}) \;\le\; \alpha \;+\; \frac{1}{\sqrt{n}} \times \max_{\theta^* \in \theta(P)} \mathbb{E}_P\left[\frac{|m(Z;\theta^*) - m(Z;\widetilde{\theta}_1)|^3}{\sigma^3(\widetilde{\theta}_1)}\right].$$

- ⋆ This implies asymptotically $1-\alpha$ coverage if the Lyapunov ratio is bounded. This boundedness holds with the consistency of $\widetilde{\theta}_1$ for many "smooth" loss functions.

- ⋆ Note that if the loss function is bounded, then one can simply use a concentration inequality instead of the CLT to get a finite-sample valid confidence set. This, for example, holds for Manski's discrete choice model.

- ⋆ None of these guarantees depend on the dimension/complexity of the parameter space Θ.

Diameter

- The diameter of the confidence set depends on the size of $\theta(P)$, the complexity of Θ, and the consistency of $\widetilde{\theta}_1$.

- Because $\widetilde{\theta}_1 \in \widehat{\mathrm{CI}}_n^{\dagger}$ and the confidence set $\widehat{\mathrm{CI}}_{n,\alpha}$ is asymptotically valid, we get

  $$\mathrm{diam}(\widehat{\mathrm{CI}}_{n,\alpha}) \;\geq\; \mathrm{diam}(\widehat{\mathrm{CI}}_n^{\dagger}) \;\geq\; \max\left\{\sup_{\theta^* \in \theta(P)} \|\theta^* - \widetilde{\theta}_1\|, \mathrm{diam}(\theta(P))\right\},$$

  with a positive probability.

- Hence, the diameter cannot converge to zero unless $\theta(P)$ is a singleton and $\widetilde{\theta}_1$ is consistent for $\theta(P)$.

- Furthermore, we also have

  $$\widehat{\theta}_2 \in \widehat{\mathrm{CI}}_n^{\dagger}, \quad \widehat{\theta}_2 := \arg\min_{\theta \in \Theta} \sum_{i \in \mathcal{I}_2} m(Z_i; \theta).$$

  The suboptimality of the M-estimator implies the suboptimality of the confidence set.

Diameter

If $\mathbb{C}(\theta') \geq \|\theta^* - \theta'\|^{1+\beta}$ for all $\theta' \in \Theta$, we get

$$\mathrm{diam}(\widehat{\mathrm{CI}}_{n,\alpha}) = O_p\left(r_n^{1/(1+\beta)} + s_n^{1/(1+\beta)}\right),$$

where

$$s_n = \mathbb{C}(\widetilde{\theta}_1) + n^{-1/2}\|m(Z;\theta(P)) - m(Z;\widetilde{\theta}_1)\|_2,$$

and r_n is such that

$$\phi_n(r_n^{2/(1+\beta)}) \asymp n^{1/2} r_n^2,$$

for

$$\phi_n(\delta) \;\geq\; \mathbb{E}\left[\sup_{\|\theta-\theta(P)\|\leq\delta} |\mathbb{G}_n[m(Z;\theta) - m(Z;\theta(P))]|\right],$$

$$\phi_n(\delta) \;\geq\; \sup_{\|\theta-\theta(P)\|\leq\delta} \|m(Z;\theta) - m(Z;\theta(P))\|_2.$$

See Theorem 8 of Takatsu and Kuchibhotla (2025, arXiv:2501.07772v3) for details.

Example: Manski's Discrete Choice Model

- ⋆ Suppose $Z_i = (X_i, Y_i)$ come from $Y_i = \mathrm{sgn}(\theta(P)^\top X_i + \xi_i)$ for some noise ξ_i that satisfies $\mathrm{Med}(\xi|X) = 0$.

- ⋆ Assume the margin condition holds:

$$\mathbb{P}_P\left(|\eta(X) - 1/2| \leq t\right) \lesssim t^{1/\beta} \quad \text{for all} \quad t < t^*,$$

  where $\eta(X) = \mathbb{P}(Y = 1|X)$.

- ⋆ If, for all $\theta \in \mathbb{S}^{d-1}$,

$$c_1\|\theta - \theta_P\|_2 \leq \mathbb{P}_P\left(\mathrm{sgn}(\theta^\top X) \neq \mathrm{sgn}(\theta_P^\top X)\right),$$

  then

$$\mathrm{diam}(\widehat{\mathrm{CI}}_{n,\alpha}) = O_p(1)\left(\frac{d\log(n/d)}{n}\right)^{1/(1+2\beta)}.$$

  This is the minimax rate.

Approach 2: Self-normalization for Z-estimation[a]

[a] Joint work with Woonyoung Chang (arXiv:2407.12278)

Z-estimation (without constraints)

- Suppose $\theta(P)$ solves the equation
$$\mathbb{E}_P[\psi(Z;\theta(P))] = 0.$$
Hence, $u^\top\psi(Z;\theta(P))$ is a mean zero random variable for any $u \in \mathbb{R}^d$.

- This implies that
$$\mathrm{CI}_{n,\alpha}(u) := \left\{\theta \in \mathbb{R}^d : \frac{|\sum_{i\in\mathcal{I}_2} u^\top\psi(Z_i;\theta)|}{\sqrt{\sum_{i\in\mathcal{I}_2}(u^\top\psi(Z_i;\theta))^2}} \le z_{\alpha/2}\right\},$$
is an asymptotically valid $(1-\alpha)$ confidence set. In fact, for all $u \in \mathbb{R}^d$ and $n \ge 1$,
$$\sup_{\theta^*\in\theta(P)} \mathbb{P}(\theta^* \notin \mathrm{CI}_{n,\alpha}(u)) \;\le\; \alpha \;+\; \frac{1}{\sqrt{n}} \times \sup_{\theta^*\in\theta(P)} \frac{\mathbb{E}_P[|u^\top\psi(Z;\theta^*)|^3]}{(\mathbb{E}_P[(u^\top\psi(Z;\theta^*)^2])^{3/2}}.$$

- This proves dimension-agnostic validity guarantee and holds for any Z-estimation problem. Note: no variance estimation, no bootstrap, no rate of convergence are needed.

Without constraints

- Although valid, this confidence set is not practically viable because it is unbounded in all but one direction. This is useful for inference for linear contrasts.

- This comes from the fact that $\mathbb{E}_P[u^\top \psi(Z;\theta)] = 0$ does not imply $\mathbb{E}_P[\psi(Z;\theta)] = 0$.

- Alternatively, vectors u that depend on θ yield bounded confidence sets. Formally,

$$\widehat{\mathrm{CI}}^*_{n,\alpha} := \left\{\theta \in \mathbb{R}^d : \frac{|\sum_{i\in\mathcal{I}_2}(\widetilde{\theta}_1 - \theta)^\top \psi(Z_i;\theta)|}{\sqrt{\sum_{i\in\mathcal{I}_2}((\widetilde{\theta}_1 - \theta)^\top \psi(Z_i;\theta))^2}} \le z_{\alpha/2}\right\},$$

  is also an asymptotically valid $(1-\alpha)$ confidence set. Here, $\widetilde{\theta}_1$ is any estimator independent of $Z_1, \ldots, Z_n$.

- The validity does not depend on the consistency of $\widetilde{\theta}_1$, but the diameter depends on it.

Without constraints

- ⋆ In the context of linear regression, this confidence set is easy to compute because it is a quadratic inequality.

- ⋆ It is clear that

$$\widetilde{\theta}_1, \widehat{\theta}_n \in \widehat{\mathrm{CI}}^*_{n,\alpha}, \quad \text{where} \quad \sum_{i \in \mathcal{I}_2} \psi(Z_i; \widehat{\theta}_n) = 0.$$

  Hence, the diameter of the confidence set cannot shrink faster than the rate of convergence of the Z-estimator.

- ⋆ Chang and Kuchibhotla (2025) show that, for linear regression,

$$\mathrm{diam}(\widehat{\mathrm{CI}}^*_{n,\alpha}) = \tilde{O}_p\left(\sqrt{d/n}\right).$$

  Similar results hold for generalized linear models, including logistic regression. For GLMs, the estimation function is

$$\psi(Z; \theta) = X(Y - \ell'(X^\top \theta)).$$

Linear Contrasts

- The advantage of these sets, in contrast to risk inversion, is in inference for linear contrasts.
- In linear regression, for any $\eta \in \mathbb{R}^d$ solving

$$\mathbb{E}[a^\top \psi(Z;\eta)] = 0 \quad \text{where} \quad a = \Sigma^{-1} c,\ \psi(Z;\eta) = X(Y - X^\top \eta),$$

  we have $c^\top \eta = c^\top \theta(P)$.
- Hence, we propose the confidence set

$$c^\top \left(\widehat{\mathrm{CI}}^*_{n,\alpha/n} \cap \mathrm{CI}_{n,\alpha}(\widetilde{\Sigma}_1^{-1} c) \right),$$

  for $c^\top \theta(P)$. This has dimension-agnostic validity and, moreover, its diameter scales as $n^{-1/2} + d/n$.

With constraints

- $\star$ The approach can be seamlessly extended to the case with constraints. Recall that if Θ is a closed convex set and $\widetilde{\theta}_1 \in \Theta$ is some initial estimator, then

$$(\widetilde{\theta}_1 - \theta(P))\mathbb{E}_P[X(Y - X^\top \theta(P))] \le 0.$$

- $\star$ Hence, a valid confidence set for $\theta(P)$ is

$$\widehat{\mathrm{CI}}^*_{n,\alpha} := \left\{ \theta \in \Theta : \frac{\sum_{i \in \mathcal{I}_2} (\widetilde{\theta}_1 - \theta)^\top \psi(Z_i; \theta)}{\sqrt{\sum_{i \in \mathcal{I}_2} ((\widetilde{\theta}_1 - \theta)^\top \psi(Z_i; \theta))^2}} \le z_{\alpha/2} \right\},$$

- $\star$ Once again, the validity is agnostic to the dimension d. The study of the diameter is in progress.
- $\star$ Construction of a confidence set for $c^\top \theta(P)$ is unresolved.

Comment: Assumptions

Set $\Sigma = \mathbb{E}[XX^\top]$ and $V = \mathbb{E}[XX^\top(Y - X^\top\theta_0)^2]$.

(LM1) There exist $q_x \geq 8, q_y, K_x, K_y \geq 1$ such that

$$\sup_{u \in \mathbb{S}^{d-1}} \mathbb{E}[|u^\top \Sigma^{-1/2} X|^{q_x}] \leq K_x^{q_x},$$

and

$$\mathbb{E}[|Y - X^\top \theta(P)|^{q_y}] \leq K_y^{q_y}.$$

Moreover, $q_{xy} := (1/q_x + 1/q_y)^{-1} \geq 4$,

(LM2) There exist positive constants $\underline{\lambda}_\Sigma$, $\overline{\lambda}_\Sigma$, $\underline{\lambda}_V$, and $\overline{\lambda}_V$ such that

$$0 < \underline{\lambda}_\Sigma \leq \lambda_{\min}(\Sigma) \leq \lambda_{\max}(\Sigma) \leq \overline{\lambda}_\Sigma < \infty$$

and

$$0 < \underline{\lambda}_V \leq \lambda_{\min}(V).$$

Comment: computation

- The proposed confidence set is

$$\widehat{\mathrm{CI}}^*_{n,\alpha} := \left\{ \theta \in \mathbb{R}^d : \frac{|\sum_{i \in \mathcal{I}_2} (\widetilde{\theta}_1 - \theta)^\top \psi(Z_i; \theta)|}{\sqrt{\sum_{i \in \mathcal{I}_2} ((\widetilde{\theta}_1 - \theta)^\top \psi(Z_i; \theta))^2}} \le z_{\alpha/2} \right\},$$

- This is analytically and computationally intractable for general ψ. Tractability can be improved using the initial estimator $\widetilde{\theta}_1$.

- Define the alternative confidence set

$$\widehat{\mathrm{CI}}^*_{n,\alpha} := \left\{ \theta \in \mathbb{R}^d : \frac{|\sum_{i \in \mathcal{I}_2} (\widetilde{\theta}_1 - \theta)^\top \psi(Z_i; \theta)|}{\sqrt{\sum_{i \in \mathcal{I}_2} ((\widetilde{\theta}_1 - \theta)^\top \psi(Z_i; \widetilde{\theta}_1))^2}} \le z_{\alpha/2} \right\},$$

- This is asymptotically valid if $\theta(P)$ is singleton, $\widetilde{\theta}_1$ is consistent for $\theta(P)$, and $\theta \mapsto \mathbb{E}[(u^\top \psi(Z; \theta))^2]$ is continuous in θ.

Conclusions

Conclusions

- ⋆ Construction of valid confidence sets can be difficult even for seemingly innocuous functionals.
- ⋆ We proposed two confidence sets for the M- and Z-estimation problems.
- ⋆ For the linear regression problem, our confidence sets are valid regardless of dimension and have a minimax diameter of $\sqrt{d/n}$.
- ⋆ Our proposal can be seamlessly extended to problems with constraints for which asymptotic limit theory is still unavailable.
- ⋆ For linear contrasts (one-dimensional functionals), our self-normalization confidence set has a diameter of order $n^{-1/2} + d/n$. In contrast, our debiasing approach yields a confidence interval with a width of $n^{-1/2}$ whenever $d = o(n^{2/3})$.
- ⋆ Characterizing the minimax width of confidence sets for linear contrasts is of interest.