Some comments on Inference under Irregularity

Arun Kumar Kuchibhotla

Carnegie Mellon University

16 October, 2026

Collaborators



Hannes Leeb (University of Vienna)

Table of contents

1. All of Inference
2. Result 1: Good intervals imply Good estimators
3. Result 2: Good estimators imply Good target
4. Examples: One, Two, . . . Aha!
5. Conclusions

Outline

All of Inference

Result 1: Good intervals imply Good estimators

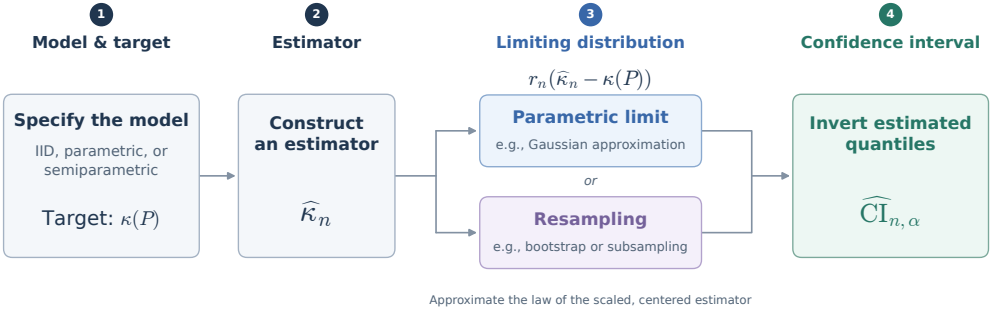
Result 2: Good estimators imply Good target

Examples: One, Two, . . . Aha!

Conclusions

All of Inference

- Inference is the fundamental contribution of statistical thinking: every estimate should be accompanied with uncertainty quantification.
- "All" of inference can be summarized as follows:



Mathematically ...

- ⋆ Suppose $\mathcal{P}$ is a collection of distributions on a measurable space. Let $\kappa : \mathcal{P} \to \mathbb{R}$ be a functional of interest.
- ⋆ Let $\widehat{\kappa}_n$ be an estimator such that for all $P_0 \in \mathcal{P}$

$$r_n(\widehat{\kappa}_n - \kappa(P_0)) \xrightarrow[P_0]{d} N(0, \sigma^2_{P_0}) \quad \text{as} \quad n \to \infty.$$

- ⋆ With $\widehat{\sigma}_n$ a consistent estimator of σ_{P_0} under P_0 (for all $P_0 \in \mathcal{P}$), the Wald interval is

$$\widehat{\mathrm{CI}}_{n,\alpha} := \left[\widehat{\kappa}_n \ \pm \ \frac{\widehat{\sigma}_n}{r_n} z_{\alpha/2}\right].$$

- ⋆ $N(0, \widehat{\sigma}_n^2/r_n^2)$ consistently estimates the distribution of $\widehat{\kappa}_n - \kappa(P_0)$ for all $P_0 \in \mathcal{P}$;
- ⋆ The Wald interval has an asymptotically exact coverage of $1 - \alpha$; and
- ⋆ The width of the Wald interval scaled by r_n is asymptotically degenerate.

... With Bootstrap

- ⋆ When the limiting distribution may not be as nice, one might use resampling techniques.
- ⋆ Suppose $X_1, \ldots, X_n$ are IID from $P_0 \in \mathcal{P}$. For $1 \leq b \leq B$, draw samples $X_1^{*b}, \ldots, X_n^{*b}$ with replacement from $\{X_1, \ldots, X_n\}$ and compute the resampling estimators

$$\widehat{\kappa}_n^{*1}, \ldots, \widehat{\kappa}_n^{*B}.$$

- ⋆ If $\widehat{q}_{\alpha/2}$ and $\widehat{q}_{1-\alpha/2}$ represent the $\alpha/2$ and $1-\alpha/2$ quantiles of these B bootstrap estimators, then the bootstrap confidence interval is given by

$$\widehat{\mathrm{CI}}_{n,\alpha}^{\mathrm{boot}} := \left[2\widehat{\kappa}_n - \widehat{q}_{1-\alpha/2},\, 2\widehat{\kappa}_n - \widehat{q}_{\alpha/2}\right].$$

- ⋆ Under usual bootstrap consistency assumption, this interval satisfies the same three properties as the Wald interval: consistency, asymptotic exactness and asymptotic degeneracy.
- ⋆ There are simple examples where bootstrap is inconsistent and several modified bootstraps or subsampling methods exist in the literature.

Desiderata for Statistical Inference

⋆ When $\mathcal{P}$ is the collection of distributions of interest and $\kappa : \mathcal{P} \to \mathbb{R}$ is a target of interest, the following are some desiderata for a confidence interval procedure $\widehat{\mathrm{CI}}_{n,\alpha}$:

⋆ Asymptotic Exactness:

$$\sup_{\alpha \in (0,1)} |\mathbb{P}_P(\kappa(P) \notin \widehat{\mathrm{CI}}_{n,\alpha}) - \alpha| \to 0 \quad \text{as} \quad n \to \infty.$$

⋆ Uniform Validity:

$$\sup_{P \in \mathcal{P}} |\mathbb{P}_P(\kappa(P) \notin \widehat{\mathrm{CI}}_{n,\alpha}) - \alpha| \to 0 \quad \text{as} \quad n \to \infty.$$

⋆ Asymptotic Degeneracy:

$$\left| \frac{\mathrm{Width}(\widehat{\mathrm{CI}}_{n,\alpha})}{a_{n,\alpha}} - 1 \right| \xrightarrow{p} 0,$$

for some non-stochastic sequence $\{a_{n,\alpha}\}$. This means that for two analysts with the same amount of information (i.e., same n), their precision is approximately the same.

Example 1: Median

- Suppose $\mathcal{P}$ is the collection of all distributions on $\mathbb{R}$ and let $\mathcal{P}_{\text{cont}}$ be the collection of all continuous distributions on $\mathbb{R}$.
- Let the CDF corresponding to $P \in \mathcal{P}$ be $F_P(x) = \mathbb{P}_P(X \leq x)$. Let the target of inference be the smallest median
$$\kappa(P) = \inf\{x : F_P(x) \geq 1/2\}.$$
- Given IID data $X_1, \ldots, X_n$ from P, a natural estimator is the sample median $\widehat{\kappa}_n$. The asymptotics of the sample median depend on the local behavior of $F_P(\cdot)$.
- Suppose
$$F_P(\kappa(P) + h) - \frac{1}{2} = A_P |h|^\gamma \text{sign}(h)(1 + o(1)) \quad \text{as} \quad h \to 0,$$
then
$$n^{1/(2\gamma)}(\widehat{\kappa}_n - \kappa(P)) \xrightarrow{d} (2A_P)^{-1/\gamma} |Z|^{1/\gamma} \text{sign}(Z) \quad \text{as} \quad n \to \infty.$$

Example 1: Median (Contd.)

- ⋆ The Wald interval cannot be implemented without the knowledge of γ.
- ⋆ The bootstrap is consistent if and only if $\gamma = 1$.
- ⋆ The subsampling is implementable if γ is known, in which case it is asymptotically exact and asymptotically degenerate. Uniform validity is unknown.
- ⋆ There is a distribution-free finite-sample valid confidence interval that is uniformly valid and asymptotically tight in $\mathcal{P}_{\text{cont}}$. Unfortunately, its width (scaled by $n^{1/(2\gamma)}$) does not converge in probability unless $\gamma = 1$.
- ⋆ Does there exist a confidence interval satisfying all the three desiderata? Note that all the statements for Wald and resampling are for sample median. Maybe there is another estimator for which all or one of them works well.

Example 2: Constrained Mean

- Suppose $\mathcal{P} = \{N(\mu, 1) : \mu \in \mathbb{R}\}$. Let $\kappa(P) = \max\{\mu, 0\}$ when $P = N(\mu, 1)$.
- Given data $X_1, \ldots, X_n \sim P = N(\mu, 1)$, a natural estimator of $\kappa(P)$ is

$$\widehat{\kappa}_n = \max\{\overline{X}_n, 0\} \quad \Rightarrow \quad n^{1/2}(\widehat{\kappa}_n - \kappa(P)) \xrightarrow{d} \begin{cases} 0, & \text{if } \mu < 0, \\ \max\{Z, 0\}, & \text{if } \mu = 0, \\ Z, & \text{if } \mu > 0 \end{cases}$$

- The Wald interval (based on $\widehat{\kappa}_n$) fails to maintain exact coverage at $\mu = 0$, the bootstrap (based on $\widehat{\kappa}_n$) is inconsistent, the subsampling (based on $\widehat{\kappa}_n$) fails to be uniformly valid.
- Does there exist a confidence interval that satisfies all the three desiderata? Note that if we have an estimator $\widetilde{\kappa}_n$ whose distribution can be estimated (uniformly) consistently, then all three conditions can be trivially satisfied.

Outline

All of Inference

Result 1: Good intervals imply Good estimators

Result 2: Good estimators imply Good target

Examples: One, Two, . . . Aha!

Conclusions

LAN Submodels

- Let $\mathcal{P}$ be a collection of distributions in a measurable space. Let $P_0 \in \mathcal{P}$.
- Let $\mathcal{B}_\delta$ be the open ball centered at zero with radius δ in $\mathbb{R}^d$.
- We say $\{P_t : t \in \mathcal{B}_\delta\}$ is an LAN submodel through P_0 if
  - $P_{t=0} = P_0$; and
  - $P_t \ll P_0$ for all $t \in \mathcal{B}_\delta$ and there exist a random vector Δ_n and a non-stochastic positive semidefinite matrix I_0 such that for each fixed $t \in \mathcal{B}_\delta$

$$\log\left(\prod_{i=1}^{n} \frac{dP_{t/\sqrt{n}}}{dP_0}(X_i)\right) \;=\; t^\top \Delta_n - \frac{1}{2} t^\top I_0 t + R_{n,t}, \qquad \text{(LAN)}$$

  with $(\Delta_n, R_{n,t})$ converging in distribution to $(N(0, I_0), 0)$ under P_0.
- Note that this is a property of the underlying collection of data-generating processes, not of the functional $\kappa(\cdot)$.

Good intervals

- Let $\kappa : \mathcal{P} \to \mathbb{R}$ be the functional of interest.
- Let $\alpha \mapsto \widehat{U}_{n,\alpha}$ be a non-increasing function of the data $X_1, \ldots, X_n$ such that

$$\sup_{\alpha \in (0,1)} \left| \mathbb{P}_{P_{t/\sqrt{n}}}(\kappa(P_{t/\sqrt{n}}) \leq \widehat{U}_{n,\alpha}) - (1-\alpha) \right| \to 0 \quad \text{as} \quad n \to \infty \quad \text{for all} \quad t \in \mathcal{B}_\delta.$$

- Moreover, for any $\varepsilon > 0$ and $\eta \in (0, 1/2)$,

$$\sup_{\alpha \in (\eta, 1-\eta)} \mathbb{P}_{P_{t/\sqrt{n}}} \left(\left| \frac{\widehat{U}_{n,\alpha} - \widehat{U}_{n,1/2}}{\mathrm{Med}(\widehat{U}_{n,\alpha} - \widehat{U}_{n,1/2})} - 1 \right| > \varepsilon \right) \to 0, \quad \text{as} \quad n \to \infty \quad \text{for all} \quad t \in \mathcal{B}_\delta.$$

- If both these conditions are satisfied, then $\widehat{\mathrm{CI}}_{n,\alpha} = (-\infty, \widehat{U}_{n,\alpha}]$ are called "good" intervals.
- The first is an asymptotic exactness condition and the second is an asymptotic degenerate width condition.

Result 1: Good Intervals imply Good Estimators

⋆ **Result 1:** Suppose $P_0 \in \mathcal{P}$ and there is an LAN submodel in $\mathcal{P}$ through P_0. Fix any non-negative sequence $a_n, n \geq 1$. Then the existence of one-sided good intervals implies the existence of a "good" estimator $\widehat{\kappa}_n = \widehat{U}_{n,1/2}$, i.e., setting

$$F_{n,t}(x) := \mathbb{P}_{P_{t/\sqrt{n}}}(a_n(\widehat{\kappa}_n - \kappa(P_{t/\sqrt{n}})) \leq x), \quad \text{for all} \quad x \in \mathbb{R}, \quad t \in \mathcal{B}_\delta,$$

as the CDF of scaled estimator, there exists a stochastic CDF $\widehat{F}_n$ such that

$$\text{Lévy}(\widehat{F}_n, F_{n,t}) \xrightarrow[P_{t/\sqrt{n}}]{p} 0 \quad \text{as} \quad n \to \infty.$$

⋆ Here a_n is arbitrary (the analyst is free to choose it). In particular, one can choose $a_n = 1$ or a_n as the rate of convergence of the estimator $\widehat{\kappa}_n$.

⋆ In simple terms, a good interval implies an estimator whose (centered) distribution can be estimated from data, locally consistently. The converse is trivially true under continuity.

⋆ Note that two-sided with confidence intervals with equi-tailed coverage give good intervals.

Outline

All of Inference

Result 1: Good intervals imply Good estimators

Result 2: Good estimators imply Good target

Examples: One, Two, . . . Aha!

Conclusions

Local Profile of a Target

- For an LAN submodel $\{P_t : t \in \mathcal{B}_\delta\}$, the local profile of the target $\kappa : \mathcal{P} \to \mathbb{R}$ at rate a_n is defined as

$$g(t) := \lim_{n\to\infty} a_n(\kappa(P_{t/\sqrt{n}}) - \kappa(P_0)) \quad \text{for} \quad t \in \mathcal{B}_\delta.$$

  Clearly, $g(0) = 0$. (One can consider a subsequence, if the limit does not exist.)

- Note that the profiles depend on the chosen sequence a_n. If $a_n \equiv 1$, then local profile is zero for continuous targets.

- If, for some sequence $a_n, n \geq 1$, there exists a $t^* \in \mathcal{B}_\delta$ such that $g(t^*) \neq 0$, then a_n represents the local minimax lower bound for estimation of $\kappa(P_{t/\sqrt{n}}), t \in \mathcal{B}_\delta$.

- Classically, the target is called pathwise differentiable if $a_n = n^{1/2}$ and $g(\cdot)$ is linear.

- It is well-known that bootstrap is consistent for Hadamard differentiable functionals (i.e., $g(\cdot)$ is linear). Some results on necessity of linearity also exist for bootstrap.

Result 2: Good estimators implies good target

- Suppose $g(\cdot)$ is the local profile of the target $\kappa : \mathcal{P} \to \mathbb{R}$ at rate a_n and suppose $g(t^*) \neq 0$ for some $t^* \in \mathcal{B}_\delta$.
- Then no estimator $\widehat{\kappa}_n$ can achieve a locally uniform rate faster than a_n for estimating $\kappa(\cdot)$.
- **Result 2:** If $a_n(\widehat{\kappa}_n - \kappa(P_{t/\sqrt{n}})) = O_p(1)$ under $P_{t/\sqrt{n}}$ and the distribution of this statistic can be locally consistently estimated, then there exists a vector $b \in \mathbb{R}^d$ such that

$$g(t) = b^\top t \quad \text{for all} \quad t \in \mathcal{B}_\delta.$$

- Combined with Result 1, we get that good intervals can only exist for functionals whose local profiles are linear.
- Alternatively, no good intervals/estimators can exist for functionals whose local profiles are non-linear.

Outline

All of Inference

Result 1: Good intervals imply Good estimators

Result 2: Good estimators imply Good target

Examples: One, Two, . . . Aha!

Conclusions

Example 1: Constrained Mean (revisited)

- Suppose $\mathcal{P} = \{N(\mu, 1) : \mu \in \mathbb{R}\}$ and let

$$\kappa(P) = \max\{\mathbb{E}_P[X], 0\} = \max\{\mu, 0\} \quad \text{if} \quad P = N(\mu, 1).$$

- Note that $\kappa(P_{t/\sqrt{n}}) = \max\{t, 0\}/\sqrt{n}$ so that

$$g(t) = n^{1/2}(\kappa(P_{t/\sqrt{n}}) - \kappa(P_0)) = \begin{cases} 0, & \text{if } t \leq 0, \\ t, & \text{if } t > 0. \end{cases}$$

- Hence, our results imply that there cannot exist an estimator $\widehat{\kappa}_n$ such that $n^{1/2}(\widehat{\kappa}_n - \kappa(P_{t/\sqrt{n}})) = O_{P_{t/\sqrt{n}}}(1)$ with a consistent estimator of its distribution.

- Additionally, there cannot exist confidence intervals that are asymptotically exact and have width scaled by $n^{1/2}$ converging in probability to a constant.

- The minimax rate scaling is important.

Example 1: Constrained Mean (Contd.)

⋆ Note that in this example, the natural estimator $\widehat{\kappa}_n := \max\{\overline{X}_n, 0\}$, satisfies

$$n^{1/2}(\widehat{\kappa}_n - \kappa(P)) = O_P(1) \quad \text{as} \quad n \to \infty \quad \text{for all} \quad P \in \mathcal{P}.$$

⋆ Therefore, the estimator

$$\widetilde{\kappa}_n = \widehat{\kappa}_n + \frac{(\log n)^{1/2}}{n^{1/2}} Z \quad \text{for} \quad Z \sim N(0,1),$$

satisfies

$$\sqrt{\frac{n}{\log n}}(\widetilde{\kappa}_n - \kappa(P)) \xrightarrow{d} Z \quad \text{for all} \quad P \in \mathcal{P}.$$

⋆ Hence, it is easy to estimate the distribution of $(n/\log n)^{1/2}(\widetilde{\kappa}_n - \kappa(P))$ can be estimated uniformly consistently.

⋆ Consequently, there are confidence intervals that are asymptotically exact with asymptotically degenerate width, albeit sub-optimal.

Example 1: Constrained Mean (Extension)

- ⋆ The constrained mean example can be generalized to many other functionals obtained via maximum or minimum.
- ⋆ Suppose $\mathcal{P}$ is a collection of distributions on $\mathbb{R}^d$ and let $X_1, \ldots, X_n$ are IID random variables from $P \in \mathcal{P}$. Let the target of interest be

$$\kappa(P) = \max_{1 \le j \le d} \mathbb{E}_P[e_j^\top X].$$

- ⋆ Take P_0 to be any distribution on $\mathbb{R}^d$ for which more than one entry of $\mathbb{E}_P[X]$ attains the maximum. Then for $J_0 = \arg\max_{1 \le j \le d} \mathbb{E}_P[e_j^\top X]$,

$$\sqrt{n}(\kappa(P_{t/\sqrt{n}}) - \kappa(P_0)) \to \max_{j \in J_0} e_j^\top t.$$

- ⋆ Similarly, for functional

$$\kappa(P) = \sup_{\lambda \in \Lambda} \mathbb{E}_P[h(X; \lambda)],$$

if at P_0, the supremum is attained at more than one $\lambda \in \Lambda$, then $\kappa(\cdot)$ can have a non-linear local profile. This includes the operator norm of a covariance matrix as an example.

Example 2: Median (revisited)

- ⋆ Suppose $\mathcal{P}$ is the collection of all distributions on $\mathbb{R}$ and let $\mathcal{P}_{\text{cont}}$ be the collection of all continuous distributions on $\mathbb{R}$.
- ⋆ Let the CDF corresponding to $P \in \mathcal{P}$ be $F_P(x) = \mathbb{P}_P(X \leq x)$. Let the target of inference be the smallest median
$$\kappa(P) = \inf\{x : F_P(x) \geq 1/2\}.$$
- ⋆ Given IID data $X_1, \ldots, X_n$ from P, a natural estimator is the sample median $\widehat{\kappa}_n$. The asymptotics of the sample median depend on the local behavior of $F_P(\cdot)$.
- ⋆ Suppose, for some sequence $\{a_n\}_{n \geq 1}$ and some strictly increasing function $\Psi(\cdot)$,
$$n^{1/2}\left(F_P\left(\kappa(P) + \frac{h}{a_n}\right) - \frac{1}{2}\right) \to \Psi(h) \quad \text{as} \quad n \to \infty, \quad \text{for each } h.$$
- ⋆ Then a_n is the local minimax rate for estimator $\kappa(P)$ and the local profile is linear if and only if $\Psi(\cdot)$ is linear.

Example 2: Median (Revisited)

- ⋆ Therefore, there cannot exist confidence intervals for the smallest median that are asymptotically exact for all continuous distribution, and simultaneously, have asymptotically degenerate width at the optimal rate, whenever $\Psi(\cdot)$ is non-linear.
- ⋆ We already know that distribution-free confidence intervals are asymptotically exact for all continuous distributions but have degenerate width if and only if $\Psi(\cdot)$ is linear.
- ⋆ Subsampling is asymptotically exact and also have asymptotically degenerate width at the optimal rate, but our results now imply that they cannot be uniformly valid.

Example 3: U-statistics

⋆ Suppose $X_1, \ldots, X_n$ are IID from a distribution P on $\mathbb{R}$. Let $h : \mathbb{R}^2 \to \mathbb{R}$ be a (symmetric) function and let the target of interest be

$$\kappa(P) := \mathbb{E}_P[h(X_1, X_2)].$$

⋆ A natural estimator of $\kappa(\cdot)$ is

$$\widehat{\kappa}_n := \frac{1}{\binom{n}{2}} \sum_{1 \le i < j \le n} h(X_i, X_j).$$

⋆ This is a U-statistics of order 2 and is a UMVUE for $\kappa(\cdot)$.

⋆ The function h is called degenerate if $\mathbb{E}_P[h(X_1, X_2)|X_1] = \kappa(P)$ almost surely. It is well-known that

$$\widehat{\kappa}_n - \kappa(P) = \begin{cases} O_P(n^{-1/2}), & \text{if } h(\cdot,\cdot) \text{ is non-degenerate under } P, \\ O_P(n^{-1}), & \text{if } h(\cdot,\cdot) \text{ is degenerate under } P. \end{cases}$$

⋆ Many U-statistics undergo such a transition between degeneracy and non-degeneracy.

Example 3: U-statistics (Contd.)

- ⋆ For example, consider
$$\kappa(P) = \mathbb{E}_P[X_1 X_2] = (\mathbb{E}_P[X])^2.$$
- ⋆ This leads to a degenerate U-statistics if $\mathbb{E}_P[X] = 0$ and a non-degenerate U-statistics if $\mathbb{E}_P[X] \neq 0$.
- ⋆ It can be proved that the local profile of $\kappa(\cdot)$ is non-linear at the minimax rate in the neighborhood of degeneracy.
- ⋆ This implies no bootstrap/resampling technique can yield a consistent estimator of the limiting distribution (locally).
- ⋆ Moreover, there cannot exist good confidence intervals for such U-functionals.

Example 4: Hellinger Projection

⋆ Let $\mathcal{F} = \{N(\theta, 1) : \theta \in \mathbb{R}\}$ be the normal parametric family. Let $\mathcal{P}$ be the collection of all distributions on $\mathbb{R}$ with a Lebesgue density, and define

$$\kappa(P) = \arg\min_{\theta \in \mathbb{R}} \mathrm{HD}(P, N(\theta, 1)) = \arg\max_{\theta \in \mathbb{R}} \int_{\mathbb{R}} \sqrt{p f_\theta} dx.$$

⋆ There must be conditions for existence and uniqueness of this minimizer. Let us consider a simple LAN submodel for which this is the case.

⋆ Set

$$P_0 = \frac{1}{2(1+e^{-2})} N(-2, 1) + \frac{e^{-2}}{(1+e^{-2})} N(0, 1) + \frac{1}{2(1+e^{-2})} N(2, 1).$$

⋆ For $t \in (-1, 1)$,

$$\frac{dP_t}{dP_0}(x) = \frac{(1 + t \tanh(x))^2}{1 + t^2 \tanh(1)} \quad \text{for all} \quad x \in \mathbb{R}.$$

⋆ This is an LAN submodel with Fisher information of $4 \tanh(1)$.

Example 4: Hellinger Projection (Contd.)

- ⋆ It can be verified that $\kappa(P_0) = 0$ and $t \mapsto \kappa(P_t)$ is a one-to-one map on $(-1, 1)$.
- ⋆ Additionally,

$$n^{1/6}(\kappa(P_{t/\sqrt{n}}) - \kappa(P_0)) \;\to\; (24)^{1/3} t^{1/3} \quad \text{as} \quad n \to \infty,$$

  which yields two interesting conclusions.

- ⋆ The local minimax rate for estimating $\kappa(P_t)$ cannot be faster than $n^{1/6}$
- ⋆ The local non-linear profile implies bootstrap/resampling and Wald intervals cannot be uniformly valid. Moreover, for no minimax rate optimal estimator, can there exist a consistent distribution estimator.
- ⋆ The local cube-root profile arises in this example because the Hellinger affinity for this parametric submodel is locally quartic at $t = 0$.

Outline

All of Inference

Result 1: Good intervals imply Good estimators

Result 2: Good estimators imply Good target

Examples: One, Two, . . . Aha!

Conclusions

Conclusions

- ⋆ Almost all traditional methods of inference start with an estimator and an estimator of the distribution of this estimator centered at the target.
- ⋆ Existing theory justifies this framework by proving its validity under regularity conditions such as "differentiability" of the functional.
- ⋆ The necessity of such differentiability assumptions is only shown for certain classes of estimators such as the plugin estimators.
- ⋆ Additionally, several other methods such as subsampling and its variants have been proposed as generic methods for valid inference under minimal regularity conditions.
- ⋆ Our results show that methods that rely on consistently estimating the limiting distribution of an estimator CANNOT succeed unless differentiability holds.
- ⋆ In particular, no clever resampling technique can overcome this deficiency. Additionally, for any method to be valid for "irregular" problems, it has to be "non-traditional."