

# Median Regularity and Honest Inference

Arun Kumar Kuchibhotla, Sivaraman Balakrishnan, and Larry Wasserman

{arunku, siva, larry}@stat.cmu.edu

Department of Statistics & Data Science, Carnegie Mellon University

## Abstract

We introduce a new notion of regularity of an estimator called median regularity. We prove that uniformly valid (honest) inference for a functional is possible if and only if there exists a median regular estimator of that functional. To our knowledge, such a notion of regularity that is *necessary* for uniformly valid inference is unavailable in the literature.

## 1 Introduction

There is a long standing distinction between pointwise and uniform inference for functionals in statistics. Pointwise confidence intervals have asymptotically correct coverage as the distribution of the data is fixed as the sample size diverges. On the other hand, uniform confidence intervals have asymptotically correct coverage even as the distribution of the data changes (i.e., triangular array) with the sample size. It is well-known that confidence intervals that only satisfy pointwise coverage validity can exhibit poor finite-sample behavior (Pötscher, 2002). This phenomenon is very well explained through the famous Hodges estimator in the normal mean example; more complicated examples can be found in post-model selection problems (Leeb and Pötscher, 2005). The distinction between pointwise and uniform inference has also been discussed in the context of resampling techniques such as bootstrap and subsampling; see Romano and Shaikh (2012); Andrews and Guggenberger (2010). Traditionally, uniformly valid confidence intervals are constructed based on a *regular* estimator (Van Der Vaart, 1991; Pfanzagl, 2000; Hirano and Porter, 2012).

Suppose we have i.i.d. data  $X_1, \dots, X_n$  from a regular parametric model  $p_{\theta_0} \in \{p_{\theta} : \theta \in \Theta\}$  for some open set  $\Theta$ . An estimator  $\hat{\tau}_n$  of  $\tau = g(\theta_0) \in \mathbb{R}$  is said to be *regular* if for some  $r_n \rightarrow \infty$  and for any  $h$ , under  $p_{\theta_0+h/r_n}$ ,

$$r_n (\hat{\tau}_n - g(\theta_0 + h/\sqrt{n})) \xrightarrow{d} L_{\theta_0}, \quad (1)$$

for some distribution  $L_{\theta_0}$  independent of  $h$  as  $n \rightarrow \infty$ . Regularity in the sense of (1) is a form of continuity of the limiting distribution (in  $\theta_0$ ). Informally, (1) says that the limiting distribution does not depend on the perturbation direction  $h$ . The theory of parametric and semiparametric efficiency is built around this notion of regularity defined in (1). Under the classical setting of parametric models that are locally asymptotically normal (LAN), the maximum likelihood estimator can be shown to be a regular estimator.

Traditionally, inference based on a regular estimator follows by estimating  $L_{\theta_0}$  (or the unknown quantities in  $L_{\theta_0}$ ) based on the observed data. Because  $L_{\theta_0}$  is the *uniform* limiting distribution of the estimator in the neighborhood, the coverage of the confidence intervals thus obtained are locally uniformly valid. The regularity notion (1) is currently the only general method of deriving (locally) uniformly valid inference (Hirano and Porter, 2012). However, a less known fact is that one does not need an estimator  $\hat{\tau}_n$  to be regular in the sense of (1) in order to achieve uniformly valid inference, i.e., the estimator need not be regular in order to construct a uniformly valid confidence interval based on it. By validity here we mean that the miscoverage probability at level  $\alpha$  is bounded above by  $\alpha$ . A simple example is the estimator  $\hat{\tau}_n = \bar{X}_n 1\{\bar{X}_n \geq 0\}$

for the functional  $\tau = \mu \mathbb{1}\{\mu \geq 0\}$ . The non-regular behavior of this estimator is well-known; see, for example, [Robins \(2004, Appendix 1\)](#). On the other hand, there are several ways to construct uniformly valid intervals based on the estimator  $\widehat{\tau}_n$  (for instance, via the HulC procedure of [Kuchibhotla et al. \(2021\)](#)).

This raises some important questions: is there a notion of regularity of the problem/estimator needed for uniformly valid inference? If so, what is that notion of regularity? We define a new notion of regularity called *median regularity* of an estimator and show that pointwise or uniformly valid inference is possible if and only if there exists an estimator which is pointwise or uniformly median regular, respectively. Furthermore, it will be clear that an estimator that is uniformly median regular leads to uniformly valid inference. We will only focus on independent and identically distributed data throughout.

## 2 Definitions and Main Result

Suppose  $X_1, \dots, X_n$  are independent and identically distributed random variables with distribution  $P$ . Let  $\theta = \theta(P)$  be a scalar functional of interest. For any estimator  $\widehat{\theta}_n$  based on  $X_1, \dots, X_n$ , define the median bias as

$$\text{Med-bias}_\theta(\widehat{\theta}_n) = \left( \frac{1}{2} - \min \left\{ \mathbb{P}(\widehat{\theta}_n \geq \theta_0), \mathbb{P}(\widehat{\theta}_n \leq \theta_0) \right\} \right)_+,$$

where  $(x)_+ = \max\{x, 0\}$ . We will write  $\text{Med-bias}_{\theta(P)}(\widehat{\theta}_n; P)$  when we want to stress that  $\theta = \theta(P)$  and the probabilities in the definition are computed when the underlying data are generated from  $P$ .

An estimator  $\widehat{\theta}_n$  is said to be median unbiased if  $\text{Med-bias}_{\theta(P)}(\widehat{\theta}_n) = 0$  ([Pfanzagl, 2011](#)). This is a finite sample property as defined and is satisfied if the probability of over-estimation and under-estimation are both at least  $1/2$ , i.e.,

$$\mathbb{P}(\widehat{\theta}_n \geq \theta(P)) \geq \frac{1}{2} \quad \text{and} \quad \mathbb{P}(\widehat{\theta}_n \leq \theta(P)) \geq \frac{1}{2}.$$

There exist several settings in the literature where finite sample median unbiased estimators exist. We only mention [Pfanzagl \(1979\)](#) that presents the most general result we know for construction of finite sample median unbiased estimators and refer the reader to Section 3 of [Kuchibhotla et al. \(2021\)](#) for further references. Here, we present here four simple examples for illustration:

1. Suppose  $X_1, \dots, X_n$  are i.i.d. random variables from  $N(\theta_0, 1)$ . Then the sample mean  $\bar{X}_n$  is median unbiased for  $\theta_0$ . The same holds true for any symmetric location family.
2. Suppose  $X_1, \dots, X_n$  are i.i.d. real valued random variables from some distribution  $P$ . Let the functional of interest be  $\theta(P)$ , the median of  $P$ . Then for any  $r \leq n/2$  the estimator

$$\widehat{\theta}_n = \begin{cases} X_{(r)}, & \text{with probability } 1/2, \\ X_{(n-r+1)}, & \text{with probability } 1/2, \end{cases}$$

is median unbiased for  $\theta(P)$ . This holds irrespective of whether the underlying distribution  $P$  has a density or not, and irrespective of whether this density is bounded away from 0 at the median or not; see [Desu and Rodine \(1969, Section 4\)](#).

3. Suppose  $X_1, \dots, X_n$  are i.i.d. random variables from  $\text{Unif}(0, \theta_0)$ . Then  $\widehat{\theta}_n = X_{(n)} - 2X_{(n-1)}$  is median unbiased for  $\theta_0$  ([Loh, 1984](#)). The MLE  $X_{(n)}$  has a median bias of  $1/2$ , the largest possible median bias for any estimator.
4. Suppose  $X_1, \dots, X_n$  are i.i.d. random variables from  $N(\mu, 1)$ . Let the functional of interest be  $\kappa(\mu) = \mu \mathbb{1}\{\mu \geq 0\}$ . Then the estimator  $\kappa(\bar{X}_n)$  is median unbiased for  $\kappa(\mu)$  ([Kuchibhotla et al., 2021, Section 3.7](#)).

Although simple median unbiased estimators exist in some parametric models, the situation is complicated in semi-/non-parametric models where one typically relies on asymptotics. We define asymptotic notions of median unbiasedness below for a family of distributions  $\mathcal{P}$ . Before describing these asymptotic notions, we remark that one can always define an estimator that takes two values  $-\infty$  and  $\infty$  with probability  $1/2$  as a median unbiased estimator of any real-valued functional  $\theta(P)$ . This is a trivial and useless estimator for practical purposes. In the following, we will ignore such a trivial estimator and call an estimator *non-trivial* if  $\mathbb{P}(|\hat{\theta}_n| < \infty) > 0$ .

**Definition 1.** *A sequence of estimators  $\{\hat{\theta}_n : n \geq 1\}$  of  $\theta(P)$  is said to be (asymptotically) median unbiased for a set of distributions  $\mathcal{P}$  if*

$$\sup_{P \in \mathcal{P}} \limsup_{n \rightarrow \infty} \text{Med-bias}_{\theta(P)}(\hat{\theta}_n; P) = 0.$$

Asymptotic median unbiasedness is a pointwise notion of median unbiasedness in the sense that we only require the median bias to converge to zero for every  $P \in \mathcal{P}$ .

**Definition 2.** *A sequence of estimators  $\{\hat{\theta}_n : n \geq 1\}$  of  $\theta(P)$  is said to be (asymptotically) median regular for a set of distributions  $\mathcal{P}$  if*

$$\limsup_{n \rightarrow \infty} \sup_{P \in \mathcal{P}} \text{Med-bias}_{\theta(P)}(\hat{\theta}_n; P) = 0.$$

Let us define a confidence interval procedure as a collection of intervals  $\{\widehat{\text{CI}}_{\alpha,n} : \alpha \in [0, 1], n \geq 1\}$  with  $\widehat{\text{CI}}_{\alpha,n}$  computable based on  $n$  observations. This means that a confidence interval procedure is just a function of the data. Below we define different notions of validity of confidence intervals. Similar to non-trivial estimators, we ignore trivial confidence intervals which return  $[-\infty, \infty]$  and call a confidence interval *non-trivial* if at least one of the end points of the confidence interval is finite almost surely.

**Definition 3.** *A confidence interval procedure  $\{\widehat{\text{CI}}_{\alpha,n} : \alpha \in [0, 1], n \geq 1\}$  for a functional  $\theta(P)$  is said to be (asymptotically) pointwise valid at level  $\gamma$  for a set of distributions  $\mathcal{P}$  if*

$$\sup_{P \in \mathcal{P}} \limsup_{n \rightarrow \infty} \mathbb{P}_P(\theta(P) \notin \widehat{\text{CI}}_{\gamma,n}) \leq \gamma.$$

**Definition 4.** *A confidence interval procedure  $\{\widehat{\text{CI}}_{\alpha,n} : \alpha \in [0, 1], n \geq 1\}$  for a functional  $\theta(P)$  is said to be (asymptotically) uniformly valid at level  $\gamma$  for a set of distributions  $\mathcal{P}$  if*

$$\limsup_{n \rightarrow \infty} \sup_{P \in \mathcal{P}} \mathbb{P}_P(\theta(P) \notin \widehat{\text{CI}}_{\gamma,n}) \leq \gamma.$$

Asymptotically uniformly valid confidence intervals are also sometimes referred to as honest intervals (Li, 1989; Pötscher, 2002).

**Definition 5.** *A confidence interval procedure  $\{\widehat{\text{CI}}_{\alpha,n} : \alpha \in [0, 1], n \geq 1\}$  for a functional  $\theta(P)$  is said to be (asymptotically) uniformly valid for a set of distributions  $\mathcal{P}$  if there exists  $\tau_n = \tau_n(\mathcal{P}) \rightarrow 0$  such that*

$$\limsup_{n \rightarrow \infty} \sup_{\gamma \in [\tau_n, 1]} \sup_{P \in \mathcal{P}} \left( \mathbb{P}_P(\theta(P) \notin \widehat{\text{CI}}_{\gamma,n}) - \gamma \right)_+ = 0.$$

The ordering of the supremums over  $\gamma \in [\tau_n, 1]$  and  $P \in \mathcal{P}$  is to make clear that  $\tau_n$  is independent of  $P$ , it is only allowed to be a function of  $\mathcal{P}$  and the sample size.

The following is the main result of this paper proved in Section 3.

**Theorem 1.** *For any set  $\mathcal{P}$  of distributions (independent of the sample size  $n$ ), the following three statements are equivalent for independent and identically distributed data:*

- (1) There exists an (asymptotically) uniformly valid non-trivial confidence interval procedure at level  $\gamma \in (0, 1)$  for  $\theta(P)$  for the set of distributions  $\mathcal{P}$ .
- (2) There exists an (asymptotically) uniformly valid non-trivial confidence interval procedure for  $\theta(P)$  for the set of distributions  $\mathcal{P}$ .
- (3) There exists a non-trivial (asymptotically) median regular estimator sequence for  $\theta(P)$  for the set of distributions  $\mathcal{P}$ .

Theorem 1 although is written for a set of distributions  $\mathcal{P}$ , there is no requirement on its cardinality; it can be an infinite set or a singleton. This flexibility allows us to use Theorem 1 to derive simple corollaries. For example, take  $\mathcal{P} = \{P_0\}$  to be a singleton with only one distribution that does not change with the sample size. Then Theorem 1 implies that asymptotically *pointwise* confidence intervals exist if and only if asymptotically median unbiased estimators exist.

**Finite sample version and impossibility results.** It is worth noting that a finite sample version of Theorem 1 can be obtained from its proof. Furthermore, one can prove that there exists a finite sample valid confidence interval at some level  $\gamma \in (0, 1)$  for  $\theta(P)$  if and only if for any  $\delta \in (0, 1/2)$ , there exists an estimator for  $\theta(P)$  with median bias bounded by  $\delta$ . This can be seen from Lemmas 1 and 2 with  $r_n = s_n = 0$  for all  $n \geq 1$ .

With this finite sample version, Theorem 1 can also be used to derive some impossibility results. For example, there exist many functionals (in non-parametric models) for which no finite sample non-trivial confidence intervals exist; the population mean is an example of such a functional (Bahadur and Savage, 1956; Pfanzagl, 1998). In such cases, the finite sample version of Theorem 1 implies that no estimator that is finite almost surely can control the median bias for such functionals; the estimator must take  $\infty$  or  $-\infty$  with non-zero probability to control the median bias at a value less than  $1/2$ . Similarly, there exist functionals for which only finite sample one-sided up/down confidence intervals exists; the number of modes of a density is an example (Donoho, 1988). In such cases again, no estimator that is finite almost surely can have median bias bounded away from  $1/2$ .

**Rate of convergence and optimality.** Unlike the classical notion of regularity (1) which explicitly indicates the rate of convergence of the estimator  $\hat{\theta}_n$ , the median regularity of an estimator does not mention/need any rate of convergence of the estimator. In fact, the estimator can be inconsistent for the true functional. This lack of reference to rate of convergence is in fact necessary, i.e., if a notion of regularity is equivalent to valid inference for all functionals, then there should be no requirement on the existence of a consistent estimator for the functionals. For example, there are functionals which are only partially identified and for such functionals no consistent estimator exists. There do exist valid confidence intervals for these functionals (Imbens and Manski, 2004; Fan and Park, 2010). Such absence of reference to rate of convergence or consistency is made possible only because we are concerned with valid coverage and not about the width of the confidence intervals. A discussion of the width of the confidence interval will typically make reference to the consistency, rate of convergence, and “optimality” of the estimator. We note in passing that optimal median regular estimators have been considered by Pfanzagl (1970) and Michel (1974).

## 3 Proof of Theorem 1

### 3.1 Intermediate Lemmas

The following key lemma shows that the existence of a valid confidence interval implies the existence of an estimator with certain bounds on its median bias.

**Lemma 1.** Fix any distribution  $P$ . Fix a known  $\gamma \in (0, 1)$ . Suppose for any sample of  $n$  independent and identically distributed observations from  $P$ , there exists a confidence interval  $\widehat{\text{CI}}_{\gamma,n}$  and  $r_n \in [0, 1)$  satisfying

$$\mathbb{P}(\theta(P) \notin \widehat{\text{CI}}_{\gamma,n}) \leq \gamma + r_n,$$

then

1. there exists an estimator  $\widehat{\theta}_n$  such that

$$\text{Med-bias}_{\theta(P)}(\widehat{\theta}_n; P) \leq \frac{\gamma + r_n}{2},$$

2. for any  $\alpha \in [0, 1]$ , there exists a confidence interval  $\widehat{\text{CI}}_{\alpha,n}$  such that

$$\mathbb{P}(\theta(P) \notin \widehat{\text{CI}}_{\alpha,n}) \leq \alpha \left(1 + 2r_{\lfloor n/B_{\alpha,\gamma/2} \rfloor}\right)^{B_{\alpha,\gamma/2}},$$

where  $B_{\alpha,\gamma/2}$  is the smallest integer  $B$  such that  $(1 - \gamma)^B + (1 + \gamma)^B \leq 2^B \alpha$ .

*Proof.* Let  $\widehat{\text{CI}}_{\gamma,n} = [\ell_n, u_n]$ . Consider the estimator

$$\widehat{\theta}_n = \begin{cases} \ell_n, & \text{with probability } 1/2, \\ u_n, & \text{with probability } 1/2. \end{cases}$$

Observe that

$$\begin{aligned} \mathbb{P}(\widehat{\theta}_n \leq \theta(P)) &= \mathbb{P}(\ell_n \leq \theta(P)) \times \frac{1}{2} + \mathbb{P}(u_n \leq \theta(P)) \times \frac{1}{2} \\ &\geq \mathbb{P}(\ell_n \leq \theta(P) \leq u_n) \times \frac{1}{2} + 0 \times \frac{1}{2} \\ &\geq (1 - \gamma - r_n)_+/2. \end{aligned}$$

Similarly,

$$\begin{aligned} \mathbb{P}(\widehat{\theta}_n \geq \theta(P)) &= \mathbb{P}(\ell_n \geq \theta(P)) \times \frac{1}{2} + \mathbb{P}(u_n \geq \theta(P)) \times \frac{1}{2} \\ &\geq 0 \times \frac{1}{2} + \mathbb{P}(\ell_n \leq \theta(P) \leq u_n) \times \frac{1}{2} \\ &\geq (1 - \gamma - r_n)_+/2. \end{aligned}$$

Therefore,

$$\min \left\{ \mathbb{P}(\widehat{\theta}_n \leq \theta(P)), \mathbb{P}(\widehat{\theta}_n \geq \theta(P)) \right\} \geq \frac{(1 - \gamma - r_n)_+}{2},$$

and hence,

$$\text{Med-bias}_{\theta(P)}(\widehat{\theta}_n; P) \leq \frac{\gamma + r_n}{2}.$$

This proves that the existence of a confidence interval implies the existence of an estimator with a specific bound on the median bias.

To construct a confidence interval at arbitrary level  $\alpha$ , Theorem 2 (and Remark 2.2) of [Kuchibhotla et al. \(2021\)](#) using the estimator  $\widehat{\theta}_n$ , its finite sample bias as approximately  $\gamma/2$ , and  $B = B_{\alpha,\gamma/2}$  yields a confidence interval  $\widehat{\text{CI}}_{\alpha,n}$  satisfying

$$\mathbb{P}(\theta(P) \notin \widehat{\text{CI}}_{\alpha,n}) \leq \alpha \left(1 + 2r_{\lfloor n/B_{\alpha,\gamma/2} \rfloor}\right)^{B_{\alpha,\gamma/2}}.$$

This completes the proof.  $\square$

**Remark.** One can construct an alternative  $\widehat{\text{CI}}_{\alpha,n}$  by considering union of independent confidence intervals. If we split the data into  $B = \lceil \log_\gamma \alpha \rceil$  batches each of size  $\lfloor n/B \rfloor$  and construct confidence intervals  $\widetilde{\text{CI}}_{\gamma, \lfloor n/B \rfloor}^{(j)}$  on each batch. Consider the confidence interval

$$\widehat{\text{CI}}_{\alpha,n} = \bigcup_{j=1}^B \widetilde{\text{CI}}_{\gamma, \lfloor n/B \rfloor}^{(j)}.$$

Observe that

$$\mathbb{P} \left( \theta(P) \notin \widehat{\text{CI}}_{\alpha,n} \right) = \prod_{j=1}^B \mathbb{P} \left( \theta(P) \notin \widetilde{\text{CI}}_{\gamma, \lfloor n/B \rfloor} \right) \leq (\gamma + r_{\lfloor n/B \rfloor})^B \leq \alpha (1 + r_{\lfloor n/B \rfloor} / \gamma)^B.$$

If  $\alpha \geq \gamma$ , then take  $B = 1$  in the above calculation; this just returns the confidence interval  $\widetilde{\text{CI}}_{\gamma,n}$ . Hence, we get  $\widehat{\text{CI}}_{\alpha,n}$  that satisfies

$$\mathbb{P} \left( \theta(P) \notin \widehat{\text{CI}}_{\alpha,n} \right) \leq \begin{cases} \alpha \left( 1 + \frac{r_{\lfloor n/\lceil \log_\gamma \alpha \rceil \rfloor}}{\gamma} \right)^{\lceil \log_\gamma \alpha \rceil}, & \text{if } \alpha \leq \gamma, \\ \alpha, & \text{if } \alpha > \gamma. \end{cases}$$

**Lemma 2.** Fix any distribution  $P$ . Fix a known  $\gamma \in (0, 1)$ . Suppose for any sample of  $n$  independent and identically distributed observations from  $P$ , there exists an estimator  $\widehat{\theta}_n$  and  $s_n \in [0, 1)$  satisfying

$$\text{Med-bias}_{\theta(P)}(\widehat{\theta}_n; P) \leq \gamma + s_n,$$

then for any  $\alpha \in (0, 1)$ , there exists a confidence interval  $\widehat{\text{CI}}_{\alpha,n}$  such that

$$\mathbb{P} \left( \theta(P) \notin \widehat{\text{CI}}_{\alpha,n} \right) \leq \alpha (1 + 2s_{\lfloor n/B_{\alpha,\gamma} \rfloor})^{B_{\alpha,\gamma}}.$$

*Proof.* The proof is exactly the same as the proof of Lemma 1(2).  $\square$

### 3.2 Proof of Theorem 1

We will prove the result by proving (1)  $\Rightarrow$  (2), (2)  $\Rightarrow$  (3), and (3)  $\Rightarrow$  (1). The aspect of non-triviality of estimators and confidence intervals follows from the proof and is not explicitly mentioned.

**(1)  $\Rightarrow$  (2):** Let  $\{\widetilde{\text{CI}}_{\gamma,n}, n \geq 1\}$  be an asymptotically uniformly valid confidence interval procedure at level  $\gamma$ , i.e.,

$$\limsup_{n \rightarrow \infty} \sup_{P \in \mathcal{P}} \left( \mathbb{P}_P(\theta(P) \notin \widetilde{\text{CI}}_{\gamma,n}) - \gamma \right)_+ = 0.$$

Define

$$r_n := \sup_{P \in \mathcal{P}} \left( \mathbb{P}_P(\theta(P) \notin \widetilde{\text{CI}}_{\gamma,n}) - \gamma \right)_+.$$

The hypothesis implies that  $r_n \rightarrow 0$ . Note that  $r_n$  is not known in practice. Lemma 1 provides a construction of a confidence interval procedure  $\{\widehat{\text{CI}}_{\alpha,n} : \alpha \in [0, 1], n \geq 1\}$  based on  $\{\widetilde{\text{CI}}_{\gamma,n} : n \geq 1\}$  that satisfies for all  $P \in \mathcal{P}$  and all  $\alpha \in [0, 1]$ :

$$\mathbb{P}_P \left( \theta(P) \notin \widehat{\text{CI}}_{\alpha,n} \right) \leq \alpha \left( 1 + 2r_{\lfloor n/B_{\alpha,\gamma/2} \rfloor} \right)^{B_{\alpha,\gamma/2}} \leq \alpha \times \exp \left( 2B_{\alpha,\gamma/2} r_{\lfloor n/B_{\alpha,\gamma/2} \rfloor} \right).$$

This implies that

$$\sup_{P \in \mathcal{P}} \left( \mathbb{P}_P(\theta(P) \notin \widehat{\text{CI}}_{\alpha,n}) - \alpha \right)_+ \leq \exp \left( 2B_{\alpha,\gamma/2} r_{\lfloor n/B_{\alpha,\gamma/2} \rfloor} \right) - 1. \quad (2)$$

Note that the right hand side of (2) only depends on  $\mathcal{P}$  through  $r_{\lfloor n/B_{\alpha,\gamma/2} \rfloor}$ . In order to prove that  $\{\widehat{\text{CI}}_{\alpha,n} : \alpha \in [0,1], n \geq 1\}$  is asymptotically uniformly valid for  $\mathcal{P}$ , it suffices to find an  $\alpha_n \rightarrow 0$  such that the right hand side of (2) converges to zero uniformly over all  $\alpha \in [\alpha_n, 1]$ . This is equivalent to proving that  $B_{\alpha,\gamma/2} r_{\lfloor n/B_{\alpha,\gamma/2} \rfloor}$  converges to zero uniformly over all  $\alpha \in [\alpha_n, 1]$ . For notational convenience, set  $\delta = \gamma/2$ . Recall that  $B_{\alpha,\delta}$  is defined as

$$B_{\alpha,\delta} = \min \{B \in \{1, 2, \dots\} : (1/2 - \delta)^B + (1/2 + \delta)^B \leq \alpha\}.$$

It is easily verified that  $B_{\alpha,\delta}$  is an increasing function of  $\alpha$  and satisfies

$$\left\lfloor \frac{\log(2/\alpha)}{\log(2)} \right\rfloor \leq B_{\alpha,\delta} \leq \left\lceil \frac{\log(2/\alpha)}{\log(2/(1+2\delta))} \right\rceil \leq \frac{2\log(2/\alpha)}{\log(2/(1+2\delta))}. \quad (3)$$

Now, define the new sequence  $\{r_k^*\}_{k \geq 1}$  based on  $\{r_k\}_{k \geq 1}$  as  $r_k^* = \sup_{m \geq k} r_m$ . It is clear that  $r_k^*$  is decreasing in  $k$  and  $r_k^* \rightarrow 0$  as  $k \rightarrow \infty$  if  $r_k \rightarrow 0$  as  $k \rightarrow \infty$ . Because  $\gamma$  and hence  $\delta$  are fixed numbers as  $n \rightarrow \infty$ ,

$$\sup_{\alpha \in [\alpha_n, 1]} \log(2/\alpha) r_{\lfloor n/B_{\alpha,\delta} \rfloor}^* \rightarrow 0 \text{ as } n \rightarrow \infty \Rightarrow \sup_{\alpha \in [\alpha_n, 1]} B_{\alpha,\delta} r_{\lfloor n/B_{\alpha,\delta} \rfloor} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Given the monotonicity properties of  $\log(2/\alpha)$  (in  $\alpha$ ) and of  $r_k^*$  in  $k$ , it is easy to prove that

$$\sup_{\alpha \in [\alpha_n, 1]} \log(2/\alpha) r_{\lfloor n/B_{\alpha,\delta} \rfloor}^* = \log(2/\alpha_n) r_{\lfloor n/B_{\alpha_n,\delta} \rfloor}^*.$$

Therefore, finding  $\alpha_n \rightarrow 0$  as  $n \rightarrow \infty$  such that the right hand side of (2) converges to zero uniformly over all  $\alpha \in [\alpha_n, 1]$  is same as finding  $\alpha_n \rightarrow 0$  such that  $\log(2/\alpha_n) r_{\lfloor n/B_{\alpha_n,\delta} \rfloor}^*$  converges to zero. Define

$$\alpha_n := \max \left\{ \frac{1}{n}, 2 \exp \left( -(r_{\lfloor n/B_{1/n,\delta} \rfloor}^*)^{-1/2} \right) \right\}.$$

From inequalities (3), it follows that  $\lfloor n/B_{1/n,\delta} \rfloor \rightarrow \infty$  as  $n \rightarrow \infty$  and hence,  $r_{\lfloor n/B_{1/n,\delta} \rfloor}^* \rightarrow 0$  as  $n \rightarrow \infty$ . This implies that the choice of  $\alpha_n$  above converges to zero as  $n \rightarrow \infty$  for any  $\delta > 0$ . Furthermore,

$$r_{\lfloor n/B_{\alpha_n,\delta} \rfloor}^* \log(2/\alpha_n) \leq r_{\lfloor n/B_{1/n,\delta} \rfloor}^* \min \left\{ \log(2n), (r_{\lfloor n/B_{1/n,\delta} \rfloor}^*)^{-1/2} \right\} \leq (r_{\lfloor n/B_{1/n,\delta} \rfloor}^*)^{1/2} \rightarrow 0,$$

as  $n \rightarrow \infty$ . Therefore, as  $n \rightarrow \infty$ ,

$$\sup_{\alpha \in [\alpha_n, 1]} \sup_{P \in \mathcal{P}} \left( \mathbb{P}_P(\theta(P) \notin \widehat{\text{CI}}_{\alpha,n}) - \alpha \right)_+ \leq \exp \left( \sup_{\alpha \in [\alpha_n, 1]} B_{\alpha,\gamma/2} r_{\lfloor n/B_{\alpha,\gamma/2} \rfloor} \right) - 1 \rightarrow 0.$$

Hence,  $\{\widehat{\text{CI}}_{\alpha,n} : \alpha \in [0,1], n \geq 1\}$  is an asymptotically uniformly valid confidence interval for the set of distributions  $\mathcal{P}$ . This completes the proof of (1) implies (2).

**(2)  $\Rightarrow$  (3):** Suppose  $\{\widetilde{\text{CI}}_{\gamma,n}, \gamma \in [0,1], n \geq 1\}$  be an asymptotically uniformly valid confidence interval procedure, i.e., there exists  $\alpha_n \rightarrow 0$  as  $n \rightarrow \infty$  such that

$$\limsup_{n \rightarrow \infty} \sup_{\alpha \in [\alpha_n, 1]} \sup_{P \in \mathcal{P}} \left( \mathbb{P}_P(\theta(P) \notin \widetilde{\text{CI}}_{\gamma,n}) - \gamma \right)_+ = 0. \quad (4)$$

Define

$$r_n := \sup_{\alpha \in [\alpha_n, 1]} \sup_{P \in \mathcal{P}} \left( \mathbb{P}_P(\theta(P) \notin \widetilde{\text{CI}}_{\gamma,n}) - \gamma \right)_+.$$

From (4),  $r_n \rightarrow 0$  as  $n \rightarrow \infty$ . Define a new confidence interval procedure  $\{\widehat{\text{CI}}_{\alpha,n} : \alpha \in [0,1], n \geq 1\}$  as

$$\widehat{\text{CI}}_{\alpha,n} := \bigcup_{\kappa \geq \alpha} \widetilde{\text{CI}}_{\kappa,n}.$$

This monotonizes the confidence intervals in terms of the confidence level. Ideally, in a family of confidence intervals, we expect that the confidence interval of confidence 90% (say) contains the confidence interval of confidence 80% (say). The initial given procedure  $\widehat{\text{CI}}_{\gamma,n}, \gamma \in [0,1]$  may not satisfy this. The defined  $\widehat{\text{CI}}_{\gamma,n}, \gamma \in [0,1]$  does satisfy this. Furthermore,

$$\mathbb{P}(\theta(P) \notin \widehat{\text{CI}}_{\gamma,n}) \leq \mathbb{P}(\theta(P) \notin \widehat{\text{CI}}_{\kappa,n}) \leq \mathbb{P}(\theta(P) \notin \widehat{\text{CI}}_{\kappa,n}) \quad \text{for all } \kappa \geq \gamma.$$

Now consider the coverage of  $\widehat{\text{CI}}_{n^{-1/2},n}$  as  $n \rightarrow \infty$ . From the monotonicity of  $\widehat{\text{CI}}_{\gamma,n}$  in  $\gamma \in [0,1]$ , it follows that

$$\mathbb{P}(\theta(P) \notin \widehat{\text{CI}}_{n^{-1/2},n}) \leq \begin{cases} \mathbb{P}(\theta(P) \notin \widehat{\text{CI}}_{\alpha_n,n}), & \text{if } \alpha_n \geq n^{-1/2}, \\ \mathbb{P}(\theta(P) \notin \widehat{\text{CI}}_{n^{1/2},n}), & \text{if } \alpha_n \leq n^{-1/2}. \end{cases} \quad (5)$$

Recall here that  $\alpha_n$  is an unknown quantity from (4). Inequality (5) follows from considering the cases  $\alpha_n \leq n^{-1/2}$  and  $\alpha_n \geq n^{-1/2}$ . From (4), we obtain

$$\mathbb{P}(\theta(P) \notin \widehat{\text{CI}}_{\alpha_n,n}) \leq \alpha_n + r_n,$$

and if  $\alpha_n \leq n^{-1/2} \leq 1$ , then

$$\mathbb{P}(\theta(P) \notin \widehat{\text{CI}}_{n^{-1/2},n}) \leq n^{-1/2} + r_n.$$

Hence, for all  $n \geq 1$ ,

$$\sup_{P \in \mathcal{P}} \mathbb{P}(\theta(P) \notin \widehat{\text{CI}}_{n^{-1/2},n}) \leq \max\{\alpha_n, n^{-1/2}\} + r_n.$$

Lemma 1 now implies that there exists an estimator  $\widehat{\theta}_n$  obtained from the confidence interval  $\widehat{\text{CI}}_{n^{-1/2},n}$  such that

$$\sup_{P \in \mathcal{P}} \text{Med-bias}_{\theta(P)}(\widehat{\theta}_n; P) \leq \frac{\max\{\alpha_n, n^{-1/2}\}}{2}.$$

Because  $\max\{\alpha_n, n^{-1/2}\}$  converges to zero as  $n \rightarrow \infty$ , we obtain an estimator sequence  $\{\widehat{\theta}_n, n \geq 1\}$  such that

$$\limsup_{n \rightarrow \infty} \sup_{P \in \mathcal{P}} \text{Med-bias}_{\theta(P)}(\widehat{\theta}_n; P) = 0.$$

This completes the proof of (2)  $\Rightarrow$  (3).

**(3)  $\Rightarrow$  (1):** Suppose  $\widehat{\theta}_n, n \geq 1$  be an estimator sequence such that

$$\limsup_{n \rightarrow \infty} \sup_{P \in \mathcal{P}} \text{Med-bias}_{\theta(P)}(\widehat{\theta}_n; P) = 0.$$

Define

$$s_n = \sup_{P \in \mathcal{P}} \text{Med-bias}_{\theta(P)}(\widehat{\theta}_n; P).$$

By the hypothesis,  $s_n \rightarrow 0$  as  $n \rightarrow \infty$ . Lemma 2 implies that there exists a confidence interval procedure  $\{\widehat{\text{CI}}_{\alpha,n} : \alpha \in [0,1], n \geq 1\}$  such that for all  $\alpha \in [0,1]$ ,

$$\sup_{P \in \mathcal{P}} \mathbb{P}(\theta(P) \notin \widehat{\text{CI}}_{\alpha,n}) \leq \alpha (1 + 2s_{\lfloor n/B_{\alpha,0} \rfloor})^{B_{\alpha,0}}.$$

Therefore,

$$\sup_{P \in \mathcal{P}} \mathbb{P}(\theta(P) \notin \widehat{\text{CI}}_{\gamma,n}) \leq \gamma (1 + 2s_{\lfloor n/B_{\gamma,0} \rfloor})^{B_{\gamma,0}}.$$

With  $\gamma \in [0,1]$  being a fixed number and  $B_{\gamma,0} = \lceil \log_2(2/\gamma) \rceil$ ,  $s_{\lfloor n/B_{\gamma,0} \rfloor} \rightarrow 0$  as  $n \rightarrow \infty$ . Hence it follows that

$$\limsup_{n \rightarrow \infty} \sup_{P \in \mathcal{P}} \mathbb{P}(\theta(P) \notin \widehat{\text{CI}}_{\gamma,n}) \leq \gamma.$$

This completes the proof of (3)  $\Rightarrow$  (1).

## 4 Summary and Conclusions

We introduced a new notion of regularity of an estimator and proved that uniformly valid inference for a functional is possible if and only if there exists a median regular estimator for that functional. The method of proof is constructive, explicitly providing an estimator from a uniformly valid confidence interval and vice versa. The confidence interval obtained from a median regular estimator is constructed via the recently proposed HulC method (Kuchibhotla et al., 2021). As a simple implication of the results of this paper we can conclude that inference for a functional is possible if and only if the HulC methodology yields valid inference with some estimator. Such an implication, to our knowledge, is not true for any other inference procedure including the bootstrap, and subsampling.

Although the method of proof is constructive, the result is mainly of theoretical interest currently. In many simple examples, we know that the uniformly valid confidence intervals or the median regular estimators obtained via the proof are sub-optimal and it is an important research direction to formalize both a useful notion of “optimality” as well as to study “optimal” constructions, i.e. procedures which yield optimal median regular estimators from (optimal) confidence intervals and vice versa.

## References

Andrews, D. W. and Guggenberger, P. (2010). Asymptotic size and a problem with subsampling and with the  $m$  out of  $n$  bootstrap. *Econometric Theory*, pages 426–468.

Bahadur, R. R. and Savage, L. J. (1956). The nonexistence of certain statistical procedures in nonparametric problems. *The Annals of Mathematical Statistics*, 27(4):1115–1122.

Desu, M. M. and Rodine, R. H. (1969). Estimation of the population median. *Scandinavian Actuarial Journal*, 1969(1-2):67–70.

Donoho, D. L. (1988). One-sided inference about functionals of a density. *The Annals of Statistics*, pages 1390–1420.

Fan, Y. and Park, S. S. (2010). Sharp bounds on the distribution of treatment effects and their statistical inference. *Econometric Theory*, 26(3):931–951.

Hirano, K. and Porter, J. R. (2012). Impossibility results for nondifferentiable functionals. *Econometrica*, 80(4):1769–1790.

Imbens, G. W. and Manski, C. F. (2004). Confidence intervals for partially identified parameters. *Econometrica*, 72(6):1845–1857.

Kuchibhotla, A. K., Balakrishnan, S., and Wasserman, L. (2021). The hulc: Confidence regions from convex hulls. *arXiv preprint arXiv:2105.14577*.

Leeb, H. and Pötscher, B. M. (2005). Model selection and inference: Facts and fiction. *Econometric Theory*, 21(1):21–59.

Li, K.-C. (1989). Honest confidence regions for nonparametric regression. *The Annals of Statistics*, 17(3):1001–1008.

Loh, W.-Y. (1984). Estimating an endpoint of a distribution with resampling methods. *The Annals of Statistics*, 12(4):1543–1550.

Michel, R. (1974). Bounds for the efficiency of approximately median unbiased estimates. *Metrika*, 21(1):205–211.

Pfanzagl, J. (1970). On the asymptotic efficiency of median unbiased estimates. *Annals of Mathematical Statistics*, 41(5):1500–1509.

Pfanzagl, J. (1979). On optimal median unbiased estimators in the presence of nuisance parameters. *Annals of Statistics*, 7(1):187–193.

Pfanzagl, J. (1998). The nonexistence of confidence sets for discontinuous functionals. *Journal of statistical planning and inference*, 75(1):9–20.

Pfanzagl, J. (2000). Are regularly estimable functionals always differentiable? *aequationes mathematicae*, 59(3):262–272.

Pfanzagl, J. (2011). *Parametric statistical theory*. Walter de Gruyter.

Pötscher, B. M. (2002). Lower risk bounds and properties of confidence sets for ill-posed estimation problems with applications to spectral density and persistence estimation, unit roots, and estimation of long memory parameters. *Econometrica*, 70(3):1035–1065.

Robins, J. M. (2004). Optimal structural nested models for optimal sequential decisions. In *Proceedings of the second seattle Symposium in Biostatistics*, pages 189–326. Springer.

Romano, J. P. and Shaikh, A. M. (2012). On the uniform asymptotic validity of subsampling and the bootstrap. *The Annals of Statistics*, 40(6):2798–2822.

Van Der Vaart, A. (1991). On differentiable functionals. *The Annals of Statistics*, pages 178–204.