Local optimization-based statistical inference

Shifeng Xiong[*]

Academy of Mathematics and Systems Science
Chinese Academy of Sciences
Beijing 100190, China
e-mail: xiong@amss.ac.cn

Abstract: This paper introduces a local optimization-based approach to test statistical hypotheses and to construct confidence intervals. This approach can be viewed as an extension of bootstrap, and yields asymptotically valid tests and confidence intervals as long as there exist consistent estimators of unknown parameters. We present simple algorithms including a neighborhood bootstrap method to implement the approach. Several examples in which theoretical analysis is not easy are presented to show the effectiveness of the proposed approach.

MSC 2010 subject classifications: Primary 62F03, 62F25, 62F40.
Keywords and phrases: Bootstrap, importance sampling, non-regular problem, resampling, space-filling design, stochastic programming.

Contents

1. Introduction

More and more complex datasets call for sophisticated statistical methods in the modern era. Compared with other fields for analyzing data such as computer science and applied mathematics, statistics can quantify the uncertainty of a phenomenon via hypothesis testing and/or interval estimation, which solidifies the unique feature of this discipline. In conventional frenquentist statistics, for testing a hypothesis or constructing a confidence interval, we need to find proper test statistic or pivotal quantity whose distribution satisfies certain properties [32]. However, this is quite difficult for many complex problems. The bootstrap method [12] relaxes the above requirement on test statistics or pivotal quantities via its ability in distribution approximation, and thus strengthens the power of conventional frequentist inference. Another advantage of bootstrap is that it provides explicit resampling-based solutions if the underlying model is well estimated. Consequently, bootstrap has been well received in statistics and other fields. The frequentist properties of bootstrap inferential procedures such as the bootstrap interval estimation can be guaranteed by the consistency of bootstrap distribution estimation [41]. This is also true for related methods like subsampling [40]. Generally speaking, it is more difficult to prove such a consistency than to derive the asymptotic distribution of the corresponding test statistic or pivotal quantity.

From the above discussion it can be seen that we have to do considerable theoretical work before claiming that the proposed method is a frequentist one. This is not easy for complex problems, and thus hampers the frequentist approach from being more applicable. In this paper we provide a very general approach based on local optimization to complement current frequentist inference. Our approach can be viewed as an extension of the classical bootstrap method, and reduces to it when the region for optimization shrinks to the center. On the theoretical aspect, the tests and confidence intervals constructed by our approach possess asymptotic frequentist properties as long as we have consistent estimators of unknown parameters. This feature indicates that we do not need to derive any (asymptotic) distribution or to prove the consistency of distribution estimation before using the proposed approach. In addition, with a proper region for optimization, the proposed approach is first order asymptotically equivalent to the bootstrap method for regular problems. On the computational aspect, our approach only requires the optimal objective value of an optimization problem over a local region, which can be reached by standard optimization techniques. We also present simple experimental design-based algorithms including a neighborhood bootstrap method to solve the optimization problem. These algorithms

are easy to implement for practitioners, and produce satisfactory results in our simulations.

The rest of this paper is organized as follows. Sections 2 and 3 introduce local optimization-based hypothesis testing and interval estimation, respectively. Their asymptotic frequentist properties are studied in Section 4. Some implementation issues are discussed in Section 5. Section 6 presents four non-regular examples including a high-dimensional problem and a nonparametric regression problem to illustrate the proposed approach. We end the paper with some discussion in Section 7.

2. Local optimization-based hypothesis testing

Let the random sample $\mathbb{X}$ be drawn from a distribution $F(\cdot, \theta)$, where θ lies in the parameter space Θ. Here Θ can be a subset of an Euclidean space or an infinite-dimensional space. We are interested in testing

$$H_0 : \ \theta \in \Theta_0 \ \leftrightarrow \ H_1 : \ \theta \in \Theta \setminus \Theta_0, \tag{2.1}$$

where Θ_0 is a close subset of Θ. Let $T = T(\mathbb{X}) \in \mathbb{R}$ be a test statistic. Suppose that T tends to take a large value when H_0 does not hold. It is known that the p-value for testing (2.1) is defined as

$$P = \sup_{\phi \in \Theta_0} \Pr(T_\phi^* \geqslant T \mid T), \tag{2.2}$$

where $T_\phi^* = T(\mathbb{X}^*)$ and $\mathbb{X}^*$ is an independent copy of $\mathbb{X}$ from $F(\cdot, \phi)$ [20]. Given a significance level $\alpha \in (0, 1)$, we will reject H_0 if $P < \alpha$. This test can strictly control the Type I error within the Neyman-Pearson framework, as shown in the following proposition [46].

Proposition 2.1. *Under H_0,*

$$\Pr(P < \alpha) \leqslant \alpha$$

Proof. Let G denote the cumulative distribution function (c.d.f.) of $-T$, i.e., $G(x) = \Pr(-T \leqslant x)$. Denote

$$G^{-1}(t) = \inf\{x : \ G(x) > t\}. \tag{2.3}$$

For $\theta \in \Theta_0$, we have

$$
\begin{aligned}
&\Pr(P < \alpha) \\
= \ & \Pr\left(\sup_{\phi \in \Theta_0} \Pr(T_\phi^* \geqslant T \mid T) < \alpha\right) \leqslant \Pr\big(\Pr(T_\theta^* \geqslant T \mid T) < \alpha\big) \quad (2.4) \\
= \ & \Pr\big(G(-T) < \alpha\big) \leqslant \Pr\big(-T < G^{-1}(\alpha)\big) \leqslant \alpha.
\end{aligned}
$$

This completes the proof. □

Proposition 2.1 is a general result, which does not require any assumption on T. From Proposition 2.1, a test is obtained by solving an stochastic optimization problem in (2.2), which can be rewritten as

$$P = \sup_{\phi \in \Theta_0} \int I(T(x) \geqslant t) dF(x, \phi), \tag{2.5}$$

where I is the indicator function and t is the realization of T. In principle, any hypothesis testing problem can be solved by this way as long as the corresponding optimization problem in (2.5) is solvable. In limited trivial cases, the problem in (2.5) has obvious solution; an example is to test simple null hypothesis

$$H_0 : \ \theta = \theta_0. \tag{2.6}$$

Sometimes T is distribution-free for all $\theta \in \Theta_0$, and then it is unnecessary to solve the optimization problem in computing P. However, except for these cases, this method faces some difficulties in computation: the stochastic optimization problem is generally very hard to solve, especially when Θ_0 is an unbounded set.

In the literature, a commonly used strategy to overcome these difficulties is based on the asymptotic distribution of the test statistic T. The optimization problem in (2.5) is often solvable when replacing the distribution of T by the asymptotic distribution. For example, with a T whose asymptotic distribution is free of unknown parameters, it is trivial to solve (2.5). For complex problems, it is often not easy to derive the asymptotic distribution, or to find such a T whose asymptotic distribution has the desirable properties. A Bayesian remedy is Meng's posterior predictive p-value [36], which averages the objective function in (2.5) over the posterior distribution of the parameter under the null hypothesis.

Here we provide a more general strategy without any requirement on the distribution of T. Suppose that H_0 holds. For the true parameter $\theta \in \Theta_0$, it suffices to obtain a p-value that controls the Type I error by optimizing the objective function in (2.2) over any set that contains θ, instead of over the whole Θ_0; see the inequality in (2.4). Consequently, we need to compute

$$P_0 = \max_{\phi \in \mathcal{N}(\theta) \cap \Theta_0} \int I(T(x) \geqslant t) dF(x, \phi), \tag{2.7}$$

where $\mathcal{N}(\theta)$ is a closed neighborhood of θ containing θ. Here "sup" in (2.5) is replaced by "max" if we assume that $\mathcal{N}(\theta) \cap \Theta_0$ is a compact subset of Θ_0 on which $\int I(T(x) \geqslant t) dF(x, \phi)$ is continuous with respect to ϕ. In practice, we use a consistent estimator $\hat{\theta}$ of θ under H_0 to replace θ in (2.7), and obtain

$$P_{\mathrm{LOT}} = \max_{\phi \in \mathcal{N}(\hat{\theta}) \cap \Theta_0} \int I(T(x) \geqslant t) dF(x, \phi). \tag{2.8}$$

If the probability of $\theta \in \mathcal{N}(\hat{\theta})$ tends to one (this is easy to achieve for consistent $\hat{\theta}$; see Section 5.1), then the test based on the p-value in (2.8) is *asymptotically valid*. We call this test *local optimization-based test* (LOT) throughout the

	How to control the Type I error	Difficulty level in implementation
Fisher's significance test	always	high
LOT	under weak conditions	moderate
Efron's bootstrap test	under strong conditions	low

paper. LOT only requires the maximum value of the objective function over a neighborhood of $\hat{\theta}$, which can be achieved by standard optimization techniques. This feature makes LOT work for many complex problems, in which it is hard to analyze the distribution of T.

When $\mathcal{N}(\hat{\theta})$ shrinks to $\hat{\theta}$, (2.8) becomes

$$P_{\mathrm{B}} = \int I(T(x) \geqslant t)dF(x, \hat{\theta}), \tag{2.9}$$

which is the p-value of the bootstrap test [11]. Therefore, LOT can be viewed as an extension of the bootstrap test. The bootstrap p-value can be easily calculated by resampling from $F(x, \hat{\theta})$, and LOT needs more computational cost. On the theoretical aspect, LOT controls the Type I error asymptotically under weak conditions, say, if $\hat{\theta}$ is a consistent estimator, whereas the bootstrap test can fail for non-regular cases where the bootstrap distribution estimator is inconsistent [1, 3]. From (2.2), (2.8), to (2.9), LOT is a bridge connecting Fisher's significance test and Efron's bootstrap test. Table 1 presents a rough description of the three tests with a given test statistic.

The above discussion mainly focuses on the case where the test statistic T is given. In current frenquentist methods including bootstrap, it is often a problem to choose a test statistic with both intuitive appeal and required distributional properties. Compared with them, LOT appears much more flexible on this point since asymptotically valid LOTs can be constructed based on any intuitively reasonable test statistic as long as we have consistent estimator of θ.

It should be emphasized that for the situations where the p-value in (2.2) is easy to compute. it is unnecessary to use complicated methods like the bootstrap and LOT. An example of such situations is to test simple hypothesis (2.6). The goal of our method is to provide reasonable solutions for complex hypothesis testing problems.

3. Local optimization-based interval estimation

The idea of approximating the p-value via local optimization can be modified to construct confidence intervals. Suppose that the parameter of interest is $\xi = \xi(\theta) \in \mathbb{R}$, and that $\hat{\xi} = \hat{\xi}(\mathbb{X})$ is an estimator of ξ. Let H_θ denote the c.d.f. of the "pivotal quantity" $\xi - \hat{\xi}$, i.e., $H_\theta(x) = \Pr(\xi - \hat{\xi} \leqslant x)$. It should be pointed out that the (asymptotic) distribution of $\xi - \hat{\xi}$ is allowed to depend on unknown parameters, and this is different from the standard definition of a pivotal quantity in textbooks. Define H_θ^{-1} as in (2.3).

Proposition 3.1. *For all $\theta \in \Theta$ and $\alpha \in (0,1)$,*

$$\Pr\left(\xi \leqslant \hat{\xi} + \sup_{\phi \in \Theta} H_\phi^{-1}(1-\alpha)\right) \geqslant 1-\alpha, \tag{3.1}$$

$$\Pr\left(\xi \geqslant \hat{\xi} + \inf_{\phi \in \Theta} H_\phi^{-1}(\alpha)\right) \geqslant 1-\alpha. \tag{3.2}$$

Proof. We have

$$\Pr\left(\xi \leqslant \hat{\xi} + \sup_{\phi \in \Theta} H_\phi^{-1}(1-\alpha)\right) \geqslant \Pr\left(\xi - \hat{\xi} \leqslant H_\theta^{-1}(1-\alpha)\right) \tag{3.3}$$

$$= H_\theta\left(H_\theta^{-1}(1-\alpha)\right) \geqslant 1-\alpha.$$

This completes the proof of (3.1), and that of (3.2) is similar. □

By Proposition 3.1, the upper and lower $1-\alpha$ confidence bounds of ξ are given by $\hat{\xi} + \sup_{\phi \in \Theta} H_\phi^{-1}(1-\alpha)$ and $\hat{\xi} + \inf_{\phi \in \Theta} H_\phi^{-1}(\alpha)$, respectively. The equal-tailed $1-\alpha$ confidence interval of ξ is $\left[\hat{\xi} + \inf_{\phi \in \Theta} H_\phi^{-1}(\alpha/2), \ \hat{\xi} + \sup_{\phi \in \Theta} H_\phi^{-1}(1-\alpha/2)\right]$. These interval limits all need to solve an optimization problem

$$\sup_{\phi \in \Theta} H_\phi^{-1}(\gamma) \quad \text{or} \quad \inf_{\phi \in \Theta} H_\phi^{-1}(\gamma)$$

for some $\gamma \in (0,1)$, which is often difficult. Like (2.7), Proposition 3.1 also holds if we take supremum over an arbitrary region containing the true value of θ; see (3.3). Suppose that $\hat{\theta}$ is a consistent estimator of θ. Under some mild conditions, we can get asymptotically valid confidence limits through solving

$$\sup_{\phi \in \mathcal{N}(\hat{\theta})} H_\phi^{-1}(\gamma) \quad \text{or} \quad \inf_{\phi \in \mathcal{N}(\hat{\theta})} H_\phi^{-1}(\gamma). \tag{3.4}$$

Specifically, the upper and lower $1 - \alpha$ confidence bounds of ξ are $\hat{\xi} + \sup_{\phi \in \mathcal{N}(\hat{\theta})} H_\phi^{-1}(1-\alpha)$ and $\hat{\xi} + \inf_{\phi \in \mathcal{N}(\hat{\theta})} H_\phi^{-1}(\alpha)$, respectively, and the equal-tailed $1 - \alpha$ confidence interval of ξ is $\left[\hat{\xi} + \inf_{\phi \in \mathcal{N}(\hat{\theta})} H_\phi^{-1}(\alpha/2), \ \hat{\xi} + \sup_{\phi \in \mathcal{N}(\hat{\theta})} H_\phi^{-1}(1 - \alpha/2)\right]$. Here "sup" (or "inf") can be replaced by "max" (or "min") if $\mathcal{N}(\hat{\theta})$ is a compact subset of Θ on which H_ϕ^{-1} is continuous with respect to ϕ. We call these confidence intervals *local optimization-based confidence intervals* (LOCIs) throughout the paper. When $\mathcal{N}(\hat{\theta})$ shrinks to $\hat{\theta}$, LOCIs become the bootstrap hybrid confidence intervals [41].

4. Asymptotic properties

This section discusses asymptotic properties of the proposed local optimization-based methods. Further results involving some computational method are deferred in the Appendix. Here we only consider one-sided LOCIs, and similar

results also hold for two-sided LOCIs and LOTs. Some notation and definitions are needed. The parameter space Θ is assumed to be a metric space with metric ρ. For $A \subset \Theta$, let $|A|$ denote $\max\{\rho(a,b) : a,b \in A\}$. For two c.d.f.'s F_1 and F_2, the Kolmogorov distance between them is defined as $d_{\mathrm{K}}(F_1, F_2) = \sup_{x \in \mathbb{R}} |F_1(x) - F_2(x)|$. We allow the neighborhood $\mathcal{N}(\cdot)$ to depend on n and denote $\mathcal{N}_n(\cdot)$ for clarity. We use "$\to_d$" to denote "converge in distribution", and let "a.s." be the abbreviation for "almost surely". As in Section 3, let H_θ denote the c.d.f. of $\xi - \hat{\xi}$. Since $\mathcal{N}_n(\hat{\theta})$ is a random set, for $\phi \in \mathcal{N}_n(\hat{\theta})$, H_ϕ is actually a random c.d.f., i.e., $H_\phi(x) = \Pr\big(\xi(\phi) - \hat{\xi}(\mathbb{X}^*) \leqslant x | \mathbb{X}\big)$, where the conditional distribution of $\mathbb{X}^*$ conditional on $\mathbb{X}$ is $F(\cdot, \phi)$.

Assumption 4.1. *As* $n \to \infty$, $\Pr\left(\theta \in \mathcal{N}_n(\hat{\theta})\right) \to 1$ *for all* $\theta \in \Theta$.

If $\hat{\theta}$ is consistent, then $\mathcal{N}_n(\hat{\theta})$ is easy to construct to satisfy Assumption 4.1; see (5.1) in Section 5.1. We can immediately have the following theorem.

Theorem 4.1. *Under Assumption 4.1, for all* $\theta \in \Theta$ *and* $\alpha \in (0,1)$,

$$\liminf_{n \to \infty} \Pr\left(\xi \leqslant \hat{\xi} + \sup_{\phi \in \mathcal{N}_n(\hat{\theta})} H_\phi^{-1}(1-\alpha)\right) \geqslant 1 - \alpha. \tag{4.1}$$

Theorem 4.1 indicates that LOCIs are asymptotically valid confidence intervals under Assumption 4.1, which is fairly weak. It is easy to find $\mathcal{N}_n(\hat{\theta})$ satisfying this assumption for consistent $\hat{\theta}$; see Section 5.1.

We next show that LOCIs are first order asymptotically equivalent to the bootstrap confidence intervals under regularity conditions. Specifically, if the bootstrap distribution estimator of $\xi - \hat{\xi}$ is consistent, then "$\geqslant$" in (4.1) can be replaced by "$=$". Several assumptions are needed.

Assumption 4.2. *As* $n \to \infty$, $|\mathcal{N}_n(\hat{\theta})| \to 0$ *(a.s.).*

Assumption 4.3. *As* $n \to \infty$, $\hat{\theta} \to \theta$ *(a.s.) for all* $\theta \in \Theta$.

Assumption 4.4. *(i) There exists a series of numbers* $a_n \to \infty$ *such that* $a_n(\xi - \hat{\xi}) \to_d K$, *where* K *is a continuous c.d.f. and is strictly increasing on its support.*
(ii) For $\phi \in \mathcal{N}_n(\hat{\theta})$, $d_{\mathrm{K}}(\bar{H}_\phi, K) \to 0$ *(a.s.), where* $\bar{H}_\phi(x) = \Pr\big(a_n[\xi(\phi) - \hat{\xi}(\mathbb{X}^*)] \leqslant x | \mathbb{X}\big)$ *and* $\mathbb{X}^*$ *is the bootstrap sample drawn from* $F(\cdot, \phi)$.

Assumption 4.4 indicates that the bootstrap distribution estimator of $a_n(\xi - \hat{\xi})$ is consistent [41]. We can use the conditional distribution of $a_n[\xi(\tilde{\theta}) - \hat{\xi}(\mathbb{X}^*)]$ conditional on $\mathbb{X}$ to approximate that of $a_n(\xi - \hat{\xi})$, and this approximation leads to asymptotically valid confidence intervals for ξ. Assumption 4.4 holds for general regular cases. We present two simple examples.

Example 4.1. *Let* X_n *be a random number from a binomial distribution* $\mathrm{BN}(n, \pi)$ *with parameter* $\pi \in (0,1)$. *Consider the pivotal quantity* $\pi - X_n/n$. *It is clear that* $\sqrt{n}(\pi - X_n/n) \to_d K$, *where* K *is the c.d.f. of* $N(0, \pi(1-\pi))$. *This*

result also holds for any strongly consistent estimator $\tilde{\pi}_n$ of π. Specifically, with $X_n^* \sim \mathrm{BN}(n, \tilde{\pi}_n)$, we can easily prove that $d_{\mathrm{K}}(\bar{H}_{\tilde{\pi}}, K) \to 0$ (a.s.) by the central limit theorem for triangle arrays, where $\bar{H}_{\tilde{\pi}}(x) = P_\pi\big(\sqrt{n}(\tilde{\pi}_n - X^*/n) \leqslant x \mid X_n\big)$, and then Assumption 4.4 holds.

Example 4.2. *Let $X_1, \ldots, X_n$ be i.i.d. random variables from a c.d.f. F with $EX_1^4 < \infty$. Here we do not assume a parametric form for F. Then the parameter space $\Theta = \{F \in \mathcal{F} : \int x^4 dF(x) < \infty\}$ is an infinite-dimensional metric space with metric d_{K}, where $\mathcal{F}$ denotes the set of all c.d.f.'s on $\mathbb{R}$. A strongly consistent estimator of F is the empirical distribution $\hat{F}(x) = \sum_{i=1}^{n} I(X_i \leqslant x)/n$. Suppose that the parameter of interest is $\mu = EX_1$. Let $\bar{X}_n$ denote the sample mean. Consider the pivotal quantity $\mu - \bar{X}_n$. First, we have $\sqrt{n}(\mu - \bar{X}_n) \to_d \Phi\big(\cdot / v(F)\big)$, where Φ is the c.d.f. of $N(0, 1)$ and $v(F) = \int \big(x - \int x dF(x)\big)^2 dF(x)$. Second, take*

$$\mathcal{N}_n(\hat{F}) = \{G \in \Theta : \ d_{\mathrm{K}}(G, \hat{F}) < 1/n^{1/3}, \ |v(G) - v(\hat{F})| < 1/n^{1/3}\}. \tag{4.2}$$

It is easy to verify Assumptions 4.1–4.3. Furthermore, for $F_n \in \mathcal{N}_n(\hat{F})$ and $X_1^*, \ldots, X_n^*$ i.i.d. from F_n, through verifying the Lindeberg condition in the central limit theorem for triangle arrays, we have that $d_{\mathrm{K}}(\bar{H}_{F_n}^{\mathrm{s}}, \Phi) \to 0$ (a.s.), where $\bar{H}_{F_n}^{\mathrm{s}}(x) = \Pr\big(\sqrt{n}(EX_1^* - \bar{X}_n^*)/v(F_n) \leqslant x \mid X_1, \ldots, X_n\big)$. Denote $\bar{H}_{F_n}(x) = \Pr\big(\sqrt{n}(EX_1^* - \bar{X}_n^*) \leqslant x \mid X_1, \ldots, X_n\big)$. By (4.2), $d_{\mathrm{K}}\big(\bar{H}_{F_n}(\cdot), \Phi\big(\cdot / v(F)\big)\big) \to 0$ (a.s.). Then Assumption 4.4 holds.

Theorem 4.2. *Under Assumptions 4.1–4.4, for all $\theta \in \Theta$ and $\alpha \in (0, 1)$,*

$$\lim_{n \to \infty} \Pr\left(\xi \leqslant \hat{\xi} + \sup_{\phi \in \mathcal{N}_n(\hat{\theta})} H_\phi^{-1}(1 - \alpha)\right) = 1 - \alpha.$$

Proof. For any n, there exists $\theta_n^* \in \mathcal{N}_n(\hat{\theta})$ such that $\sup_{\phi \in \mathcal{N}_n(\hat{\theta})} \bar{H}_\phi^{-1}(1 - \alpha) < \bar{H}_{\theta_n^*}^{-1}(1 - \alpha) + 1/n$. Under Assumptions 4.2 and 4.3, $\theta_n^* \to \theta_0$ (a.s.). Therefore, by Assumption 4.4, $\bar{H}_{\theta_n^*}^{-1}(1 - \alpha) \to K^{-1}(1 - \alpha)$ (a.s.). We have

$$
\begin{aligned}
\Pr\left(\xi \leqslant \hat{\xi} + \sup_{\phi \in \mathcal{N}_n(\hat{\theta})} H_\phi^{-1}(1 - \alpha)\right) &= \Pr\left(a_n(\xi - \hat{\xi}) \leqslant \sup_{\phi \in \mathcal{N}_n(\hat{\theta})} \bar{H}_\phi^{-1}(1 - \alpha)\right) \\
&\leqslant \Pr\left(a_n(\xi - \hat{\xi}) \leqslant \bar{H}_{\theta_n^*}^{-1}(1 - \alpha) + 1/n\right) \\
&= \Pr\left(a_n(\xi - \hat{\xi}) \leqslant K^{-1}(1 - \alpha) + o(1)\right) \\
&= \bar{H}_\theta\left(K^{-1}(1 - \alpha) + o(1)\right) \to 1 - \alpha,
\end{aligned}
$$

where $\bar{H}_\theta$ is the c.d.f. of $a_n(\xi - \hat{\xi})$ that converges to K by Assumption 4.4 (i). Combining this result with Theorem 4.1, we complete the proof. $\square$

When applying bootstrap to a specific problem, we need to verified Assumptions 4.3 and 4.4 to guarantee its frequentist properties. Theorems 4.1 and 4.2

indicate that we do not need to do such theoretical work when using LOCI. With a proper $\mathcal{N}_n(\hat{\theta})$, LOCI possesses both the basic frequentist property in (4.1) and a potential bonus: it enjoys the same first order frequentist property as the bootstrap method when the two assumptions hold (although we may not know this). It can be expected that, under much stronger conditions, LOCI has some high-order asymptotic properties like bootstrap [25]. We do not discuss this in the paper since it is difficult to verify such conditions for complex problems.

5. Implementation

This section discusses how to implement LOT and LOCI. We focus on the cases where Θ is a subset of an Euclidean space. Therefore, it suffices to solve finite-dimensional optimization problems in LOT and LOCI. For some problems with infinite-dimensional parameter spaces, LOT or LOCI is still available through rational simplification; see Section 6.4.

5.1. Specification of $\mathcal{N}(\hat{\theta})$

The first issue is to determine the neighborhood $\mathcal{N}(\hat{\theta})$ in (2.8) and (3.4) over which we solve the optimization problem. Suppose that the dimension of Θ is q and $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_q)'$ is a consistent estimator of $\theta = (\theta_1, \ldots, \theta_q)'$. The basic principle is to select $\mathcal{N}(\hat{\theta})$ satisfying Assumption 4.1. A simple choice of $\mathcal{N}(\hat{\theta})$ is

$$\left[\hat{\theta}_1 - \delta, \ \hat{\theta}_1 + \delta\right] \times \cdots \times \left[\hat{\theta}_q - \delta, \ \hat{\theta}_q + \delta\right] \tag{5.1}$$

for some small constant $\delta > 0$. If we know further the convergence rate of $\hat{\theta}$, then the second principle is to select $\mathcal{N}(\hat{\theta})$ satisfying Assumption 4.2. By Theorem 4.2, this selection can make the local optimization-based method asymptotically equivalent to bootstrap if the bootstrap distribution estimator is consistent. For example, with $\|\hat{\theta} - \theta\| = O_p(1/\sqrt{n})$, a selection of $\mathcal{N}(\hat{\theta})$ simultaneously satisfying Assumptions 4.1 and 4.2 is

$$\left[\hat{\theta}_1 - \delta \log(n)/\sqrt{n}, \ \hat{\theta}_1 + \delta \log(n)/\sqrt{n}\right] \times \cdots \times \left[\hat{\theta}_q - \delta \log(n)/\sqrt{n}, \ \hat{\theta}_q + \delta \log(n)/\sqrt{n}\right] \tag{5.2}$$

for some constant $\delta > 0$. The constant δ in (5.1) or (5.2) can be specified empirically. For complex problems, the convergence rate of $\hat{\theta}$ is difficult to exactly know. We will see in Section 6 that, LOT or LOCI has good finite-sample performance even with a simple $\mathcal{N}(\hat{\theta})$ like in (5.1) that only satisfies Assumption 4.1.

It seems more reasonable if the variances of $\hat{\theta}_j$'s are used to construct $\mathcal{N}(\hat{\theta})$. When the variance estimators are not straightforward, the jackknife, bootstrap [41], or even Bayesian methods can be used to estimate the variances. However, such methods will add extra theoretical and computational work, and there are still some constants, which need to be specified empirically, in the final form of $\mathcal{N}(\hat{\theta})$. Therefore, we suggest using the variance estimators only when they are straightforward.

5.2. Importance sampling-based approach

Suppose that $F(\cdot, \theta)$ has a probability density function (p.d.f.) $f(\cdot, \theta)$ with respect to a σ-finite measure ν, and that $\{f(\cdot, \theta) : \phi \in \Theta_0\}$ has a common support. We use an *importance sampling-based approach* to solve the stochastic optimization problem in (2.8). First we approximate the objective function in (2.8) by importance sampling. Note that

$$u(\phi) = \int I(T(x) \geqslant t) \frac{f(x, \phi)}{f(x, \hat{\theta})} f(x, \hat{\theta}) d\nu(x) = E\left\{ I(T(\mathbb{X}^*) \geqslant t) \frac{f(\mathbb{X}^*, \phi)}{f(\mathbb{X}^*, \hat{\theta})} \right\},$$

where $\mathbb{X}^* \sim f(\cdot, \hat{\theta})$. According to the sample averaging approximation method in stochastic optimization [42], we compute the p-value as

$$P_{\mathrm{IS}} = \max_{\phi \in \mathcal{N}(\hat{\theta}) \cap \Theta_0} \hat{u}(\phi), \tag{5.3}$$

where

$$\hat{u}(\phi) = \frac{1}{M} \sum_{m=1}^{M} \left\{ I(T(\mathbb{X}_m^*) \geqslant t) \frac{f(\mathbb{X}_m^*, \phi)}{f(\mathbb{X}_m^*, \hat{\theta})} \right\}, \tag{5.4}$$

is the approximation of $u(\theta)$ based on $\mathbb{X}_1^*, \ldots, \mathbb{X}_M^*$ i.i.d. from $f(\cdot, \hat{\theta})$ with the Monte Carlo sample size M. With sufficiently large M, P_{IS} can be arbitrarily close to P in (2.8). There are many available iterative algorithms for solving the deterministic optimization problem in (5.3) such as the interior point method [5].

We can also use an experimental design-based method to approximate the p-value in (5.3). Take L points $\phi_1, \ldots, \phi_L$ uniformly spaced over $\mathcal{N}(\hat{\theta}) \cap \Theta_0$, and then compute

$$P_{\mathrm{IS-D}} = \max \left\{ \hat{u}(\hat{\theta}), \hat{u}(\phi_1), \ldots, \hat{u}(\phi_L) \right\}, \tag{5.5}$$

where $\hat{u}$ is defined in (5.4). We call these points *trial points* throughout this paper, which can be constructed from so-called *space-filling designs* in experimental design; see Section 5.4. Since $\mathcal{N}(\hat{\theta})$ is a small neighborhood, $P_{\mathrm{IS-D}}$ often performs well with a moderate L. The design-based method is very easy to implement, and is suitable for those who are not familiar with optimization methods. More sophisticated space-filling design-based optimization method can be found in [18].

For LOCI, we have the following importance sampling-based method to compute the interval limits when $F(\cdot, \theta)$ has a p.d.f. $f(\cdot, \theta)$ and $\{f(\cdot, \theta) : \theta \in \Theta\}$ has a common support. Here we only consider the computation of upper limits, i.e., the first optimization problem in (3.4). Let $\varphi = H_\phi^{-1}(\gamma)$ and $S(\phi, \varphi) = H_\phi(\varphi)$. Suppose that H_ϕ is continuous and strictly increasing on its support for $\phi \in \mathcal{N}(\hat{\theta})$. The problem (3.4) is equivalent to the constrained optimization problem

$$\max_{\phi \in \mathcal{N}(\hat{\theta})} \varphi \quad \text{subject to } S(\phi, \varphi) = \gamma. \tag{5.6}$$

For q-dimensional space Θ, the problem optimizes $q+1$ variables. Similar to the importance sampling-based sample averaging approximation method in (5.4), we use an approximation of S,

$$\hat{S}(\phi, \varphi) = \frac{1}{M} \sum_{m=1}^{M} \left\{ I\big(\xi(\phi) - \hat{\xi}(\mathbb{X}_m^*) \leqslant \varphi\big) \frac{f(\mathbb{X}_m^*, \phi)}{f(\mathbb{X}_m^*, \hat{\theta})} \right\},$$

where $\mathbb{X}_1^*, \ldots, \mathbb{X}_M^*$ are i.i.d. from $f(\cdot, \hat{\theta})$ with the Monte Carlo sample size M. The solution to

$$\max_{\phi \in \mathcal{N}(\hat{\theta})} \varphi \quad \text{subject to } \hat{S}(\phi, \varphi) = \gamma \tag{5.7}$$

can be used to approximate that to (5.6). Note that $\hat{S}(\phi, \varphi)$ may not equal γ exactly in (5.7). In practice we handle an equivalent problem

$$\max_{\phi \in \mathcal{N}(\hat{\theta})} \varphi \quad \text{subject to } \hat{S}(\phi, \varphi) \leqslant \gamma \tag{5.8}$$

instead of (5.7). A design-based method similar to (5.5) can also be used to solve (5.8). Since (5.8) has not straightforward solution even for a given $\phi \in \mathcal{N}(\hat{\theta})$, we do not recommend such a method. A more simple and general method for computing LOCIs is to directly compute the quantiles of H_ϕ for a given ϕ. This method will be discussed in the next subsection.

5.3. Neighborhood bootstrap

This subsection discusses a general method, called *neighborhood bootstrap*, to implement LOT and LOCI. This method still works for the cases where the importance sampling-based approach in Section 5.2 fails. We first consider LOT. Like the design-based p-value in (5.5), take L trial points $\phi_1, \ldots, \phi_L$ uniformly spaced over $\mathcal{N}(\hat{\theta}) \cap \Theta_0$. The difference from (5.5) is that the neighborhood bootstrap method directly approximates the objective value in (2.8) by the Monte Carlo method. Specifically, for each ϕ_l, $l = 0, 1, \ldots, L$, generate $\mathbb{X}_{l,1}^*, \ldots, \mathbb{X}_{l,M}^*$ i.i.d. from $F(\cdot, \phi_l)$, where $\phi_0 = \hat{\theta}$. Then the p-value in (2.8) can be approximated by

$$P_{\text{NB}} = \max_{l=0,\ldots,L} \left\{ \frac{1}{M} \sum_{m=1}^{M} I\big(T(\mathbb{X}_{l,m}^*) \geqslant t\big) \right\}.$$

For LOCI, we still consider the computation of upper limits in (3.4). With $\{\phi_1, \ldots, \phi_L\}$ uniformly spaced over $\mathcal{N}(\hat{\theta})$, take bootstrap sample $\mathbb{X}_{l,1}^*, \ldots, \mathbb{X}_{l,M}^*$ i.i.d. from $F(\cdot, \phi_l)$ for $l = 0, 1, \ldots, L$. Let $\widehat{H_{\phi_l}^{-1}(\gamma)}$ denote the sample γ-quantile of $\xi(\phi_l) - \hat{\xi}(\mathbb{X}_{l,1}^*), \ldots, \xi(\phi_l) - \hat{\xi}(\mathbb{X}_{l,M}^*)$. Consequently, $\sup_{\phi \in \mathcal{N}(\hat{\theta})} H_\phi^{-1}(\gamma)$ can be approximated by

$$\max_{l=0,\ldots,L} \left\{ \widehat{H_{\phi_l}^{-1}(\gamma)} \right\}. \tag{5.9}$$

Neighborhood bootstrap is a very general method. In principle, it can be applied to infinite-dimensional parameter spaces if there are well-defined space-filling designs for such spaces. Another advantage of neighborhood bootstrap is its easy implement, especially for computing LOCIs. For LOT, neighborhood bootstrap is slightly more time-consuming than the importance sampling-based approach.

5.4. Selection of trial points

The design-based p-value in (5.5) and the neighborhood bootstrap method in Section 5.3 both need L trial points $\phi_1, \ldots, \phi_L$ uniformly spaced over $\mathcal{N}(\hat{\theta})$. This subsection presents some discussion on the design of these points. Usually $\mathcal{N}(\hat{\theta})$ is selected as a q-dimensional hypercube like (5.1) or (5.2). Specifically, suppose that $\mathcal{N}(\hat{\theta}) = [L_1, U_1] \times \cdots \times [L_q, U_q]$. For $\psi_i = (\psi_{i1}, \ldots, \psi_{iq})' \in [0,1]^q$, $i = 1, \ldots, L$, let $\phi_{ij} = L_j + \psi_{ij}(U_j - L_j)$, $i = 1, \ldots, L$, $j = 1, \ldots, q$, and we have $\phi_i = (\phi_{i1}, \ldots, \phi_{iq})' \in \mathcal{N}(\hat{\theta})$ for $i = 1, \ldots, L$. Therefore, it suffices to consider the design of $\psi_1, \ldots, \psi_L$ in $[0,1]^q$, called *initial design* in the following. As mentioned in Section 5.2, the initial design can be constructed from space-filling designs in $[0,1]^q$. Such designs include grids, Latin hypercube designs (LHD) [35], and uniform designs [19], among others. A simple grid is given by

$$\left\{ \frac{1}{2U}, \ldots, \frac{2U-1}{2U} \right\} \times \cdots \times \left\{ \frac{1}{2U}, \ldots, \frac{2U-1}{2U} \right\}, \qquad (5.10)$$

where U is a positive integer. There are $L = U^q$ points in the grid, and this leads to unaffordable computations for large q. LHDs are easy to construct for any L and q. The LHD is spaced uniformly in each dimension, and its space-filling properties over the whole $[0,1]^q$ can be improved by iterative algorithms [38]. There are functions for generating LHDs in both `MATLAB` and `R`.

Note that in fact we need to design $\phi_1, \ldots, \phi_L$ in $\mathcal{N}(\hat{\theta}) \cap \Theta_0$ for LOT or in $\mathcal{N}(\hat{\theta}) \cap \Theta$ for LOCI. For irregular or constrained parameter spaces, this problem becomes complicated. A feasible solution is to design more points in $\mathcal{N}(\hat{\theta})$ and then to keep those in the intersection.

6. Illustrative examples

This section presents four examples to illustrate LOT and LOCI, in which the (asymptotic) distributions of the test statistics or pivotal quantities are non-regular or unclear.

6.1. Interval estimation for the maximum cell probability of the multinomial distribution

Let $(X_{n1}, \ldots, X_{nk})'$ be the cell frequencies from a multinomial distribution, $\mathrm{MN}_k(n; \pi)$, where $\sum_{i=1}^{k} X_{ni} = n$, with the parameter $\pi = (\pi_1, \ldots, \pi_k)'$, $\pi_i >$

0, $i = 1, \ldots, k$, and $\sum_{i=1}^{k} \pi_i = 1$. We consider interval estimation for $\pi_{\max} = \max\{\pi_1, \ldots, \pi_k\}$. This problem is related to some real applications including the diversity of ecological populations [39] and favorable numbers on a roulette wheel [14], and has been studied by Gelfand et al. [22], Glaz and Sison [24], and Xiong and Li [49], among others.

The maximum likelihood estimator (MLE) of π is $(X_{n1}/n, \ldots, X_{nk}/n)'$, and that of $\pi_{\max}$ is $\max_{1 \leqslant i \leqslant k} X_{ni}/n$. To avoid extreme values in the estimators, we use the Bayesian estimator $\hat{\pi} = (\hat{\pi}_1, \ldots, \hat{\pi}_k)' = \big((X_{n1} + 1/2)/(n + k/2), \ldots, (X_{nk} + 1/2)/(n + k/2)\big)'$ from the Jeffrey prior [23]. The corresponding estimator of $\pi_{\max}$ is $\hat{\pi}_{\max} = \max_{1 \leqslant i \leqslant k} \hat{\pi}_i$, whose asymptotic properties are the same as the MLE. Xiong and Li [49] showed that, when the numbers in $\{i = 1, \ldots, k : \pi_i = \pi_{\max}\}$ are more than one, $\hat{\pi}_{\max}$ is not asymptotically normal and the corresponding bootstrap distribution estimator is inconsistent. A remedy is to use m-out-of-n bootstrap [2]. This method takes bootstrap sample $(X_{m1}^*, \ldots, X_{mk}^*)'$ from $\mathrm{MN}_k(m; \hat{\pi})$ with $m = o(n)$, and then approximates the distribution of $\sqrt{n}(\hat{\pi}_{\max} - \pi_{\max})$ by its bootstrap analogue $\sqrt{m}(\hat{\pi}_{\max}^* - \hat{\pi}_{\max})$, where $\hat{\pi}_{\max}^* = \max_{1 \leqslant i \leqslant k} \big\{(X_{mi}^* + 1/2)/(m + k/2)\big\}$. Xiong and Li [50] proved that this approximation is consistent, and thus results in asymptotically valid confidence intervals for $\pi_{\max}$.

The LOCI of $\pi_{\max}$ can be easily constructed by the neighborhood bootstrap method in (5.9), where the pivotal quantity is $\pi_{\max} - \hat{\pi}_{\max}$. We next conduct a simulation study to compare the LOCI with the ordinary bootstrap and m-out-of-n bootstrap methods. Here we focus on two-sided $1 - \alpha$ confidence intervals with $\alpha = 0.05$. In our simulation study, k is fixed as 5, and $n = 30$ and 60 are considered. We use six vectors of cell probabilities; see Table 2. In the m-out-of-n bootstrap method, m is set as the integer part of $2\sqrt{n}$. The neighborhood $\mathcal{N}(\hat{\pi})$ is

$$\big[\hat{\pi}_1 - \delta \log(n)/\sqrt{n}, \ \hat{\pi}_1 + \delta \log(n)/\sqrt{n}\big]$$
$$\times \cdots \times \big[\hat{\pi}_k - \delta \log(n)/\sqrt{n}, \ \hat{\pi}_k + \delta \log(n)/\sqrt{n}\big]$$

where two values, 0.1 and 0.5, of δ are used. it is clear that $\mathcal{N}(\hat{\pi})$ satisfies Assumptions 4.1 and 4.2. We use two grids in (5.10) to design the trial points with $U = 3$ for $\delta = 0.1$ and $U = 5$ for $\delta = 0.5$. Note that there is a constraint $\sum_{i=1}^{k} \pi_i = 1$ in the parameter space. There are 51 and 101 trial points in the two grids, respectively. The bootstrap sample size is 5000 in all the above methods.

We repeat 5000 times to compute the coverage rates (CRs), mean lengths (MLs), and standard deviations of lengths (SDLs) of the confidence intervals. The simulation results are shown in Table 2. We can see that the bootstrap interval usually has low CR. For dispersed π, the m-out-of-n bootstrap method lacks efficiency with longer ML, whereas two LOCIs perform better. As expected, the LOCI with $\delta = 0.5$ is more conservative than that with $\delta = 0.1$. In summary, it can be concluded that the LOCI is at least comparable to the m-out-of-n bootstrap interval.

TABLE 2
Simulation results in Section 6.1

$\pi = (0.7, 0.075, 0.075, 0.075, 0.075)'$

	$n = 30$			$n = 60$		
	CR	ML	SDL	CR	ML	SDL
Bootstrap	0.927	0.299	0.023	0.932	0.222	0.014
Bootstrap $(m < n)$	0.920	0.389	0.054	0.945	0.377	0.037
LOCI $(\delta = 0.1)$	0.950	0.325	0.021	0.940	0.228	0.013
LOCI $(\delta = 0.5)$	0.961	0.345	0.020	0.954	0.236	0.012

$\pi = (0.5, 0.15, 0.15, 0.1, 0.1)'$

	$n = 30$			$n = 60$		
	CR	ML	SDL	CR	ML	SDL
Bootstrap	0.846	0.297	0.040	0.912	0.237	0.011
Bootstrap $(m < n)$	0.881	0.368	0.065	0.955	0.362	0.038
LOCI $(\delta = 0.1)$	0.897	0.321	0.033	0.931	0.244	0.008
LOCI $(\delta = 0.5)$	0.967	0.350	0.018	0.967	0.259	0.008

$\pi = (0.3, 0.175, 0.175, 0.175, 0.175)'$

	$n = 30$			$n = 60$		
	CR	ML	SDL	CR	ML	SDL
Bootstrap	0.738	0.175	0.075	0.702	0.147	0.058
Bootstrap $(m < n)$	0.748	0.187	0.093	0.722	0.164	0.090
LOCI $(\delta = 0.1)$	0.939	0.210	0.075	0.832	0.172	0.054
LOCI $(\delta = 0.5)$	0.991	0.296	0.046	0.990	0.217	0.032

$\pi = (0.3, 0.3, 0.2, 0.1, 0.1)'$

	$n = 30$			$n = 60$		
	CR	ML	SDL	CR	ML	SDL
Bootstrap	0.893	0.207	0.064	0.909	0.168	0.035
Bootstrap $(m < n)$	0.913	0.234	0.081	0.955	0.210	0.065
LOCI $(\delta = 0.1)$	0.954	0.248	0.064	0.944	0.205	0.036
LOCI $(\delta = 0.5)$	0.979	0.327	0.036	0.966	0.241	0.023

$\pi = (0.24, 0.24, 0.24, 0.24, 0.04)'$

	$n = 30$			$n = 60$		
	CR	ML	SDL	CR	ML	SDL
Bootstrap	0.935	0.170	0.065	0.924	0.134	0.041
Bootstrap $(m < n)$	0.956	0.184	0.074	0.986	0.149	0.057
LOCI $(\delta = 0.1)$	0.943	0.210	0.064	0.949	0.174	0.042
LOCI $(\delta = 0.5)$	0.976	0.305	0.032	0.970	0.220	0.024

$\pi = (0.2, 0.2, 0.2, 0.2, 0.2)'$

	$n = 30$			$n = 60$		
	CR	ML	SDL	CR	ML	SDL
Bootstrap	0.906	0.135	0.063	0.946	0.095	0.049
Bootstrap $(m < n)$	0.982	0.136	0.074	0.996	0.088	0.059
LOCI $(\delta = 0.1)$	0.950	0.175	0.064	0.963	0.127	0.047
LOCI $(\delta = 0.5)$	0.937	0.280	0.038	0.963	0.195	0.028

6.2. Interval estimation for the location parameter of the three-parameter Weibull distribution

The Weibull distribution is widely used in many fields such as survival analysis [8] and reliability [37]. Let $X_1, \ldots, X_n$ be i.i.d. observations from the Weibull distribution $\mathrm{Wbl}(a, b, \tau)$, whose p.d.f. is

$$f(x; a, b, \tau) = \frac{b}{a} \left(\frac{x - \tau}{a} \right)^{b-1} \exp \left[- \left(\frac{x - \tau}{a} \right)^b \right] \tag{6.1}$$

for $x > \tau$, $a > 0$, $b > 0$, and $\tau \in \mathbb{R}$. The parameters a, b, and τ are known as the scale, shape, and location parameters, respectively. If τ is known, then the likelihood-based inference for the parameters is straightforward [37]. With an unknown τ, the standard method faces difficulties since the distributions have not a common support [4]. Estimation for the parameters of the three-parameter Weibull distribution is still an active topic in recent years, and many estimators have been proposed [33, 7, 44]. Since the (asymptotic) distributions of these estimators are difficult to derive, there is limited results on interval estimation for the parameters.

This subsection constructs LOCIs for τ based on the maximum product of spacings (MPS) estimation [6]. Obviously our method is also applicable for other parameters. The MPS estimators $\hat{a}$, $\hat{b}$, and $\hat{\tau}$ are constructed by maximizing

$$S(a, b, \tau) = \prod_{i=1}^{n+1} \int_{X_{(i-1)}}^{X_{(i)}} f(x; a, b, \tau) dx,$$

where $X_{(1)} \leqslant \cdots \leqslant X_{(n)}$ are order statistics, $X_{(0)} = \tau$, and $X_{(n+1)} = \infty$. For all a, b, and τ, the MPS estimators are consistent [6]. Furthermore, for $b > 2$, they have the same asymptotic distributions as the MLEs; if $0 < b \leqslant 2$, then $\hat{a} - a = O_p(1/\sqrt{n})$, $\hat{b} - b = O_p(1/\sqrt{n})$, and $\hat{\tau} - \tau = O_p(1/n^{1/b})$. It is not straightforward to construct confidence intervals of τ by the asymptotic properties of $\hat{\tau}$ since b is unknown. Furthermore, the validity of the corresponding bootstrap confidence interval is unclear.

We use neighborhood bootstrap to construct two-sided $1 - \alpha$ confidence intervals of τ, and conduct a simulation study to evaluate their performance. The pivotal quantity is $\tau - \hat{\tau}$. The initial design is the grid in (5.10) with $U = 3$ that corresponds to $L = 27$. Since the results are sensitive to the value of b, we set the neighborhood $\mathcal{N}(\hat{a}, \hat{b}, \hat{\tau})$ as

$$\left[\hat{a} - \delta_n, \ \hat{a} + \delta_n \right] \times \left[\hat{b} - \delta_n, \ \hat{b} + \delta_n \right] \times \left[\hat{\tau} - \delta_n, \ \hat{\tau} + \delta_n \right],$$

where $\delta_n = 4 \exp \left(- (1/\hat{b})^5 \right) \log(n)/\sqrt{n}$. It is clear that $\mathcal{N}(\hat{a}, \hat{b}, \hat{\tau})$ satisfies Assumptions 4.1 and 4.2 for all a, b, and τ by the asymptotic properties of the MPSs. For $\tau = 1$, two values of n, and several combinations of (a, b), the simulation results based on 1000 repetitions are reported in Table 3 with $\alpha = 0.05$. The bootstrap sample sizes used in the bootstrap interval and LOCI are both

TABLE 3
Simulation results in Section 6.2

	$a = 0.5,\ b = 0.5$					
	$n = 10$			$n = 20$		
	CR	ML	SDL	CR	ML	SDL
Bootstrap	0.970	0.112	0.075	0.945	0.023	0.017
LOCI	0.979	0.129	0.190	0.946	0.024	0.020
	$a = 2.5,\ b = 0.5$					
	$n = 10$			$n = 20$		
	CR	ML	SDL	CR	ML	SDL
Bootstrap	0.946	0.336	0.223	0.940	0.072	0.057
LOCI	0.946	0.343	0.234	0.940	0.073	0.060
	$a = 0.5,\ b = 1.5$					
	$n = 10$			$n = 20$		
	CR	ML	SDL	CR	ML	SDL
Bootstrap	0.410	0.072	0.042	0.230	0.024	0.012
LOCI	0.949	1.013	0.546	0.982	0.789	0.316
	$a = 2.5,\ b = 1.5$					
	$n = 10$			$n = 20$		
	CR	ML	SDL	CR	ML	SDL
Bootstrap	0.210	0.240	0.133	0.098	0.055	0.029
LOCI	0.876	1.081	0.644	0.921	1.166	0.542
	$a = 0.5,\ b = 2.5$					
	$n = 10$			$n = 20$		
	CR	ML	SDL	CR	ML	SDL
Bootstrap	0.289	0.104	0.042	0.121	0.038	0.014
LOCI	0.958	1.454	0.472	0.972	1.246	0.189
	$a = 2.5,\ b = 2.5$					
	$n = 10$			$n = 20$		
	CR	ML	SDL	CR	ML	SDL
Bootstrap	0.014	0.221	0.068	0.010	0.038	0.020
LOCI	0.881	1.392	0.671	0.950	1.614	0.379

1000. We can see that, for $b = 0.5$, the CR of the bootstrap interval is satisfactory, and the LOCI has similar performance to it with slightly longer ML. For larger b, the bootstrap interval performs poorly, and the LOCI is much better in terms of CR.

6.3. *Testing whether all the coefficients in the high-dimensional regression are nonnegative*

High-dimensional data analysis that deals with models where the number of parameters is larger than the sample size is a very active research area in recent

years, We consider the regression model

$$y = X\beta + \varepsilon, \tag{6.2}$$

where $X = (x_{ij})$ is the $n \times p$ regression matrix, $y = (y_1, \ldots, y_n)'$ is the response vector, $\beta = (\beta_1, \ldots, \beta_p)'$ is the vector of regression coefficients and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)'$ is a vector of i.i.d. normal random errors with zero mean and finite variance σ^2. Let p_0 denote the number in $\{j = 1, \ldots, p : \beta_j \neq 0\}$. For $p \gg n$, we make the sparsity assumption of $p_0 \ll n$. Many methods have been proposed to estimate the sparse β in (6.2) such as the lasso [45], the smoothly clipped absolute deviation method [16], and the minimax concave penalty method [51]. Under the assumption that all the coefficients are known to be nonnegative, Efron et al. [13] introduced a nonnegative lasso method to estimate β, which solves

$$\min_{\beta} \|y - X\beta\|^2 + \lambda \sum_{j=1}^{p} \beta_j \quad \text{subject to } \beta_j \geqslant 0, \ j = 1, \ldots, p, \tag{6.3}$$

where $\lambda > 0$ is a tuning parameter. Applications of this method can be found in [21, 47]. In this subsection we use the data to test whether the assumption in the nonnegative lasso method is reasonable, i.e., test the following hypotheses

$$H_0 : \beta_j \geqslant 0, \ j = 1, \ldots, p \ \leftrightarrow \ H_1 : \text{There exists } j_0 \in \{1, \ldots, p\} \text{ such that } \beta_{j_0} < 0. \tag{6.4}$$

In classical $n > p$ settings, the problem to test (6.4) has been discussed by the likelihood ratio test [43]. However, this method cannot be dirrectly extended to the high-dimensional case since the MLEs perform very poorly for such a case. Here we borrow the idea of the generalized likelihood ratio test in nonparametric statistics [17], and construct the test statistic

$$T = \frac{\|y - X\hat{\beta}_{H_0}\|^2}{\|y - X\hat{\beta}_{H_1}\|^2},$$

where $\hat{\beta}_{H_0}$ and $\hat{\beta}_{H_1}$ are the estimators of β under H_0 and H_1, respectively. A natural choice is to use the nonnegative lasso estimator in (6.3) and the lasso estimator as $\hat{\beta}_{H_0}$ and $\hat{\beta}_{H_1}$, respectively. Since the distribution of T under H_0 is unclear, we use LOT to test (6.4).

First of all we need to estimate all the unknown parameters under H_0. Wu, Yang, and Liu [47] showed that the nonnegative lasso estimator in (6.3) is consistent under H_0. By [15], a consistent estimator of σ^2 is $\hat{\sigma}^2 = \|y - X\hat{\beta}_{LS}\|^2/n$, where $\hat{\beta}_{LS}$ is the ordinary least squares estimator of β under the submodel selected by the nonnegative lasso. Since p is large, the neighborhood $\mathcal{N}(\hat{\beta}_{H_0}, \hat{\sigma}^2)$ should be selected elaborately to avoid high-dimensional optimization. We select $\mathcal{N}(\hat{\beta}_{H_0}, \hat{\sigma}^2)$ as

$$\mathcal{N}(\hat{\beta}_{H_0,1}) \times \cdots \times \mathcal{N}(\hat{\beta}_{H_0,p}) \times \mathcal{N}(\hat{\sigma}^2), \tag{6.5}$$

where $\hat{\beta}_{H_0,j}$'s are components of $\hat{\beta}_{H_0}$, $\mathcal{N}(\hat{\beta}_{H_0,j}) = \{0\}$ for $\hat{\beta}_{H_0,j} = 0$ and $\mathcal{N}(\hat{\beta}_{H_0,j}) = [\hat{\beta}_{H_0,j} - \delta\hat{\sigma}, \ \hat{\beta}_{H_0,j} + \delta\hat{\sigma}]$ otherwise, $\mathcal{N}(\hat{\sigma}^2) = [\hat{\sigma}^2 - \delta, \ \hat{\sigma}^2 + \delta]$, and

$\delta > 0$ is a constant. By the importance sampling-based approach in Section 5.2, the p-value of the LOT for (6.4) is given by (5.3). Note that the asymptotic results in Section 4 cannot be directly applied for diverging p. However, it is not hard to show that, if H_0 holds, then $\Pr\big((\beta, \sigma^2) \in \mathcal{N}(\hat{\beta}_{H_0}, \hat{\sigma}^2)\big) \to 1$ as $n \to \infty$ under regularity conditions by selection consistency properties of the nonnegative lasso [47]. Therefore, similar to Theorem 4.1, the asymptotic frequentist property of the LOT can be guaranteed.

We conduct a simulation study to compare the above LOT and the bootstrap test whose p-value is given in (2.9). In the simulation the rows of X in (6.2) are i.i.d. from a multivariate normal distribution $N(0, \Sigma)$ whose covariance matrix $\Sigma = (\sigma_{ij})_{p \times p}$ has entries $\sigma_{ii} = 1$, $i = 1, \ldots, p$ and $\sigma_{ij} = 0.1$, $i \neq j$. The random errors $\varepsilon_1, \ldots, \varepsilon_n$ i.i.d. $\sim N(0, 1)$. We use three configurations of n and p, $(n, p) = (20, 40)$, $(n, p) = (40, 80)$, and $(n, p) = (60, 120)$. We take the tuning parameter $\lambda = 4\sqrt{\log(p)/n}$ in the lasso and nonnegative lasso estimator recommended by Wu, Yang, and Liu [47]. In the LOT, δ is set as 0.03 in (6.5), and we compute the p-value in (5.5) with 30 trial points. Here the initial design of the trial points is an LHD, whose dimension is the number of non-zero $\hat{\beta}_{H_0, j}$'s; see (6.5). In the two methods, the bootstrap sample sizes are both 2000. The significance levels $\alpha = 0.05$ and $\alpha = 0.1$ are considered.

TABLE 4
Type I errors in Section 6.3

			$n = 20,\ p = 40$					
	$\alpha = 0.05$				$\alpha = 0.1$			
	(i)	(ii)	(iii)	(iv)	(i)	(ii)	(iii)	(iv)
Bootstrap	0.084	0.094	0.096	0.092	0.176	0.199	0.186	0.192
LOT	0.048	0.056	0.056	0.050	0.099	0.119	0.108	0.122
			$n = 40,\ p = 80$					
	$\alpha = 0.05$				$\alpha = 0.1$			
	(i)	(ii)	(iii)	(iv)	(i)	(ii)	(iii)	(iv)
Bootstrap	0.134	0.172	0.160	0.168	0.248	0.302	0.282	0.308
LOT	0.052	0.062	0.062	0.060	0.110	0.126	0.116	0.126
			$n = 60,\ p = 120$					
	$\alpha = 0.05$				$\alpha = 0.1$			
	(i)	(ii)	(iii)	(iv)	(i)	(ii)	(iii)	(iv)
Bootstrap	0.206	0.216	0.194	0.239	0.372	0.378	0.362	0.376
LOT	0.066	0.054	0.060	0.051	0.128	0.126	0.100	0.136

Four vectors of the coefficients under H_0 are used: (i) $\beta_1 = \cdots = \beta_p = 0$; (ii) $\beta_1 = 2$ and $\beta_j = 0$ for other j; (iii) $\beta_1 = \beta_2 = 2$ and $\beta_j = 0$ for other j; (iv) $\beta_1 = \beta_2 = \beta_3 = 2$ and $\beta_j = 0$ for other j. To compute the power, we consider $\beta_1 = 2$, $\beta_2 = c < 0$, and $\beta_j = 0$ for other j. For each model, we simulate 2000 data sets, and report the Type I errors and powers in Table 4 and Figure 1, respectively. It can be seen that the bootstrap test cannot control the Type I error well, and that the LOT has reasonable performance in terms of
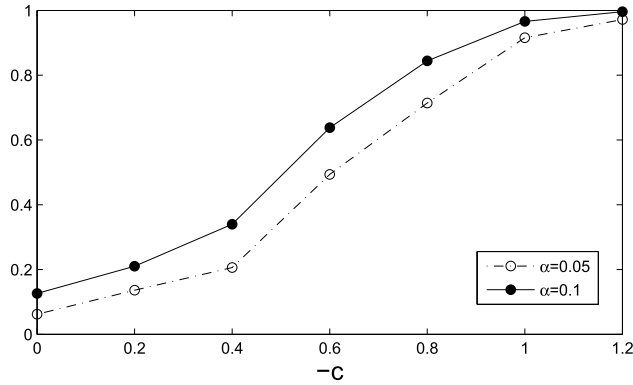
FIG 1. *Powers of the LOT in Section 6.3 ($n = 40, p = 80$).*

the Type I error and power. The power performance of the LOT is similar for other parameter configurations.

6.4. Interval estimation for the minimum of an unknown function

Consider the nonparametric regression model

$$y = r(x) + \varepsilon, \tag{6.6}$$

where r is a continuous function defined on $[0, 1]$ and $\varepsilon \sim N(0, \sigma^2)$ is the random error. For given $x_1, \ldots, x_n \in [0, 1]$, the responses are denoted by $y_1, \ldots, y_n$, respectively, where $y_i = r(x_i) + \varepsilon_i$ and $\varepsilon_1, \ldots, \varepsilon_n$ are independent. Assume that r has a unique minimum ξ in $[0, 1]$, i.e., $r(\xi) < r(x)$ for all $x \in [0, 1]$ with $x \neq \xi$. We are interested in constructing confidence intervals of $\xi = \xi(r)$. Without the random error, some related problems have been discussed in the literature [10, 11]. However, to the best of the author's knowledge, there is no result on interval estimation for ξ in the regression setting.

In model (6.6), the unknown parameter r lies in an infinite-dimensional space. We shall show that, with a fixed design for $x_1, \ldots, x_n$, the problem of construction confidence intervals for ξ can be simplified to a finite-dimensional problem, and thus can be solved by the approaches in Section 5. Here we only focus on the upper $1 - \alpha$ confidence interval for $\alpha \in (0, 1)$. Let $\hat{r}$ and $\hat{\sigma}$ be estimators of r and σ. We use $\hat{\xi}(\mathbb{X}) = \arg\min_{x \in [0,1]} \hat{r}(x)$ as an estimator of ξ with $\mathbb{X} = (y_1, \ldots, y_n)'$, and consider the pivotal quantity $\xi(r) - \hat{\xi}(\mathbb{X})$ with the c.d.f. $H_{(r,\sigma)}(x) = \Pr(\xi(r) - \hat{\xi}(\mathbb{X}) \leqslant x)$. For $a \in [0, 1]$, $b_1, \ldots, b_n \in \mathbb{R}$, and $c > 0$, let $\mathbb{X}^* = (y_1^*, \ldots, y_n^*)$ denote the set of independent random variables $y_i^* \sim N(b_i, c^2)$ for $i = 1, \ldots, n$. Let $\theta = (a, b_1, \ldots, b_n, c)'$ and $\tilde{H}_\theta(x) = \Pr(a - \hat{\xi}(\mathbb{X}^*) \leqslant x)$. Denote by $\mathcal{N}_n(\hat{\theta}) \subset \mathbb{R}^{n+2}$ a neighborhood of $\hat{\theta} = (\hat{\xi}, \hat{r}(x_1), \ldots, \hat{r}(x_n), \hat{\sigma})'$. Since $H_{(r,\sigma)}(x) = \tilde{H}_{(\xi(r), r(x_1), \ldots, r(x_n), \sigma)'}(x)$, the following proposition is straightforward.
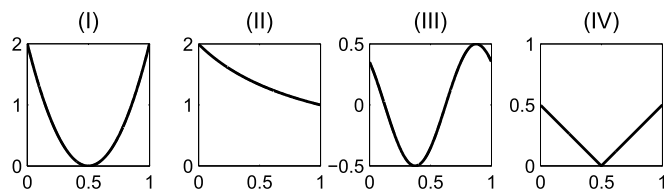
Proposition 6.1. *For all r and σ, if* $\Pr\big((\xi, r(x_1), \ldots, r(x_n), \sigma)' \in \mathcal{N}_n(\hat{\theta})\big) \to 1$, *then*

$$\liminf_{n \to \infty} \Pr\left(\xi \leqslant \hat{\xi} + \sup_{\phi \in \mathcal{N}_n(\hat{\theta})} \tilde{H}_\phi^{-1}(1-\alpha)\right) \geqslant 1 - \alpha.$$

TABLE 5
Simulation results in Section 6.4

		(I)				
		$n = 20$			$n = 30$	
	CR	ML	SDL	CR	ML	SDL
Bootstrap	0.635	0.264	0.066	0.643	0.243	0.060
LOCI	0.933	0.487	0.062	0.948	0.457	0.055
		(II)				
		$n = 20$			$n = 30$	
	CR	ML	SDL	CR	ML	SDL
Bootstrap	0.749	0.189	0.176	0.766	0.161	0.159
LOCI	0.935	0.281	0.220	0.950	0.251	0.199
		(III)				
		$n = 20$			$n = 30$	
	CR	ML	SDL	CR	ML	SDL
Bootstrap	0.710	0.272	0.091	0.656	0.233	0.069
LOCI	0.973	0.573	0.127	0.963	0.490	0.095
		(IV)				
		$n = 20$			$n = 30$	
	CR	ML	SDL	CR	ML	SDL
Bootstrap	0.721	0.486	0.203	0.721	0.446	0.175
LOCI	0.927	0.794	0.192	0.957	0.787	0.161

By Proposition 6.1, we can obtain LOCIs of ξ which have the asymptotic frenquentist properties through optimizing the quantiles of $\tilde{H}_\phi$ over a local region. In the following, let $\hat{r}$ be the Nadaraya-Watson estimator with kernel function K and bandwidth h [28]. Under regularity conditions, $\sup_{x \in [0,1]} |\hat{r}(x) - r(x)| \to 0$ in probability [27], which implies that $\hat{\xi}$ is a consistent estimator of ξ. Additionally, a consistent estimator $\hat{\sigma}^2$ of σ^2 can be given from the residual sum of

squares of $\hat{r}$. A choice of $\mathcal{N}_n(\hat{\theta})$ satisfying the condition in Proposition 6.1 is

$$
\begin{aligned}
\left[\hat{\xi} - \delta\hat{\sigma},\ \hat{\xi} + \delta\hat{\sigma}\right] &\times \left[\hat{r}(x_1) - \delta\hat{\sigma},\ \hat{r}(x_1) + \delta\hat{\sigma}\right] \\
&\times \cdots \times \left[\hat{r}(x_n) - \delta\hat{\sigma},\ \hat{r}(x_n) + \delta\hat{\sigma}\right] \times \left[\hat{\sigma} - \delta,\ \hat{\sigma} + \delta\right], \quad (6.7)
\end{aligned}
$$

where $\delta > 0$ is a constant.

We next conduct a simulation study to compare the bootstrap two-sided $1 - \alpha$ confidence intervals and LOCIs with $\alpha = 0.05$. Four regression functions in (6.6) are considered:

$$
\begin{aligned}
&\text{(I)}: \ r(x) = 2(2x - 1)^2; \quad \text{(II)}: \ r(x) = 2/(x + 1); \\
&\text{(III)}: \ r(x) = \sin(2\pi x + 3\pi/4)/2; \quad \text{(IV)}: \ r(x) = |x - 1/2|;
\end{aligned}
$$

see Figure 2. For these functions, the values of ξ are $1/2$, 1, $3/8$, and $1/2$, respectively. We fix $\sigma^2 = 1/4$, and $x_i = (2i - 1)/(2n)$ for $i = 1, \ldots, n$. The kernel function K in $\hat{r}$ is the Epanechnikov kernel, and the bandwidth h is set as $n^{-1/5}/5$. In LOCIs, we use $\delta = 0.25$ in (6.7), and take 60-run LHDs as the initial designs of trial points for implementing neighborhood bootstrap. The bootstrap sample size is 5000. Based on 5000 repetitions, we report the simulation results in Table 5. It can be seen that the bootstrap method performs poorly in terms of CR, and that the LOCI is much better for all the cases.

7. Discussion

In this paper we have introduced the local optimization-based inference including LOT and LOCI. The main advantage of our approach is that, unlike current frequentist approach, it does not require hard work in deriving (asymptotic) distributions since its asymptotic frequentist properties hold as long as we have consistent estimators of the unknown parameters. The implementation of our approach is based on standard computational methods such as importance sampling and Monte Carlo, which are easy to master for practitioners. Local optimization-based inference can be viewed as an extended bootstrap that complements current frequentist inference. It can fast provide frequentist solutions to complex problems in practice, and has broadly potential applications. Illustrative examples have shown these to some extent. Although local optimization-based inference does not overshoot for regular problems (see Theorem 4.2), it is more suitable for non-regular problems in which the theoretical derivation is difficult. The local optimization-based method may not be very accurate or efficient, but can act as *the last method* when we cannot derive other frequentist methods.

We give a further discussion on the specification of the neighborhood $\mathcal{N}(\hat{\theta})$ here. Generally speaking, the choice of $\mathcal{N}(\hat{\theta})$ is flexible; see Section 6. In real applications, for a dataset with fixed sample size n, it is not hard to find a proper $\mathcal{N}(\hat{\theta})$ that guarantees that LOT or LOCI has satisfactory performance via empirical evaluations. Besides the methods in Section 5.1, we can also use informative priors, if any, to inform the construction of $\mathcal{N}(\hat{\theta})$. This provides a

way to associate our approach with Bayesian statistics, and is valuable to study in the future. A related problem to the specification of $\mathcal{N}(\hat{\theta})$ is that it is difficult to get the exact solution or to know how close an approximate one to it even for a small $\mathcal{N}(\hat{\theta})$. This problem is not very serious in practice since our terminal is inference instead of optimization. Simulation results in Section 6 show that the design-based approximation with a moderate L yields satisfactory finite-sample performance of LOT and LOCI even for high-dimensional $\mathcal{N}(\hat{\theta})$. In fact, when bootstrap gives aggressive results, local optimization-based inference can always improve its performance, even with a relatively poor optimization algorithm.

The local optimization-based inference uses computer-intensive methods to solve complex hypothesis testing and interval estimation problems. A disadvantage of such methods is high computational cost. We can replace the Monte Carlo method in the implementation with LHD sampling or quasi Monte Carlo to improve the computational efficiency [29]. Iterative algorithms such as stochastic approximation [31] are also available to solve the stochastic programming problem in (2.8). In addition, the local optimization-based method is cumbersome when θ is a high-dimensional parameter due to the curse of dimensionality, and is hardly feasible in a nonparametric setting due to the difficulty of infinite-dimensional optimization. It is a possible future topic to study how to overcome these difficulties. Combinations of our method with other inferential methods [26, 48, 34] are also valuable to study in the future.

Appendix A: Asymptotic properties of the design-based algorithm

As mentioned in Section 5, $\max_{l=1,\ldots,L_n} H_{\phi_l}^{-1}(1-\alpha)$ can be used to approximate the upper limit $\sup_{\phi\in\mathcal{N}_n(\hat{\theta})} H_\phi^{-1}(1-\alpha)$ in LOCI, where $\{\phi_1,\ldots,\phi_{L_n}\}$ is a dense subset of $\mathcal{N}_n(\hat{\theta})$. We next prove frequentist properties of this approximation. These results are less important in practice since we can obtain an approximation as accurate as possible with a powerful computer. We place them here because they may be still of interest in theory.

Assumption A.1. *Let $\{a_n\}$ be a series of positive numbers. Denote $\bar{H}_\phi(x) = \Pr\big(a_n[\xi(\phi)-\hat{\xi}(\mathbb{X}^*)] \leqslant x\,|\,\mathbb{X}\big)$, where $\mathbb{X}^*$ is drawn from $F(\cdot,\phi)$ given $\mathbb{X}$. As $n\to\infty$,*

$$\max_{\phi\in\mathcal{N}_n(\hat{\theta})} \min_{l=1,\ldots,L_n} \big|\bar{H}_\phi^{-1}(1-\alpha) - \bar{H}_{\phi_l}^{-1}(1-\alpha)\big| = o_p(1).$$

Assumption A.2. *As $n\to\infty$ and $\delta\to 0$, $\bar{H}_\theta\big(\bar{H}_\theta^{-1}(1-\alpha)-\delta\big)\to 1-\alpha$.*

Note that the limits of $\bar{H}_{\varphi_1}$ and $\bar{H}_{\varphi_2}$ can be different for $\varphi_1,\ \varphi_2 \in \mathcal{N}_n(\hat{\theta})$. Assumption A.1 requires that $\{\phi_1,\ldots,\phi_{L_n}\}$ should be dense enough so that any value of $\bar{H}_\phi^{-1}(1-\alpha)$ can be approximated accurately by some element in $\{\bar{H}_{\phi_l}^{-1}(1-\alpha)\}_{l=1,\ldots,L_n}$. Assumption A.1 holds under Assumptions 4.2–4.4, and relates to some space-filling criterion [30]. Assumption A.2 says that $\bar{H}_\theta$ is asymptotically continuous at $\bar{H}_\theta^{-1}(1-\alpha)$. Under Assumption 4.4 (i), Assumption A.2 holds.

Theorem A.1. *Under Assumptions [4.1], [A.1], and [A.2],*

$$\liminf_{n\to\infty} \Pr\left(\xi \leqslant \hat\xi + \max_{l=1,\ldots,L_n} H_{\phi_l}^{-1}(1-\alpha)\right) \geqslant 1-\alpha.$$

Proof. For any n, there exists $\theta_n^* \in \mathcal{N}_n(\hat\theta)$ such that $\sup_{\phi\in\mathcal{N}_n(\hat\theta)} \bar{H}_\phi^{-1}(1-\alpha) < \bar{H}_{\theta_n^*}^{-1}(1-\alpha) + 1/n$. Denote $l^* = \arg\min_{l=1,\ldots,L_n} \left|\bar{H}_{\phi_l}^{-1}(1-\alpha) - \bar{H}_{\theta_n^*}^{-1}(1-\alpha)\right|$. We have $\max_{l=1,\ldots,L_n} \bar{H}_{\phi_l}^{-1}(1-\alpha) \geqslant \bar{H}_{\phi_{l^*}}^{-1}(1-\alpha) \geqslant \bar{H}_{\theta_n^*}^{-1}(1-\alpha) - |\bar{H}_{\phi_{l^*}}^{-1}(1-\alpha) - \bar{H}_{\theta_n^*}^{-1}(1-\alpha)|$. Therefore,

$$\begin{aligned}
&\Pr\left(\xi \leqslant \hat\xi + \max_{l=1,\ldots,L_n} H_{\phi_l}^{-1}(1-\alpha)\right) \\
&= \Pr\left(a_n(\xi-\hat\xi) \leqslant \max_{l=1,\ldots,L_n} \bar{H}_{\phi_l}^{-1}(1-\alpha)\right) \\
&\geqslant \Pr\left(a_n(\xi-\hat\xi) \leqslant \bar{H}_{\theta_n^*}^{-1}(1-\alpha) - |\bar{H}_{\phi_{l^*}}^{-1}(1-\alpha) - \bar{H}_{\theta_n^*}^{-1}(1-\alpha)|\right) \\
&\geqslant \Pr\left(a_n(\xi-\hat\xi) \leqslant \sup_{\phi\in\mathcal{N}_n(\hat\theta)} \bar{H}_\phi^{-1}(1-\alpha) + o_p(1) - 1/n\right) \\
&\geqslant \Pr\left(a_n(\xi-\hat\xi) \leqslant \bar{H}_\theta^{-1}(1-\alpha) + o_p(1)\right) - \Pr\left(\theta \in \mathcal{N}_n(\hat\theta)\right) \\
&= 1-\alpha + o(1) - \Pr\left(\theta \in \mathcal{N}_n(\hat\theta)\right) \to 1-\alpha,
\end{aligned}$$

which completes the proof. □

The following theorem is straightforward.

Theorem A.2. *Under Assumptions [4.1]–[4.4], [A.1], and [A.2], for all $\theta \in \Theta$ and $\alpha \in (0,1)$,*

$$\lim_{n\to\infty} \Pr\left(\xi \leqslant \hat\xi + \max_{l=1,\ldots,L_n} H_{\phi_l}^{-1}(1-\alpha)\right) = 1-\alpha.$$

Acknowledgements

References

[1] ANDREWS, D. W. (2000). Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica* **68** 399–405. MR1748009

[2] BICKEL, P. J., GÖTZE F, and VAN ZWET, W. R. (1997). Resampling fewer than n observations gains, losses, and remedies for losses. *Statistica Sinica* **7** 1–31. MR1441142

[3] BICKEL, P. J. and REN, J-J. (2001). The bootstrap in hypothesis testing, Lecture Notes-Monograph Series. *State of the Art in Probability and Statistics* **36** 91–112. MR1836556

[4] BLISCHKE, W. R. (1974). On nonregular estimation. II. Estimation of the location parameter of the gamma and Weibull distributions. *Communications in Statistics* **3** 1109–1129. MR0431478

[5] BOYD, S. and VANDENBERGHE, L. (2004). *Convex Optimization.* Cambridge University Press, Cambridge. MR2061575

[6] CHENG, R. C. H. and AMIN, N. A. K. (1983). Estimating parameters in continuous univariate distributions with a shifted origin. *Journal of the Royal Statistical Society, Ser. B* **45** 394–403. MR0737651

[7] COUSINEAU, D. (2009). Fitting the three-parameter Weibull distribution: Review and evaluation of existing and new methods. *IEEE Transactions on Dielectrics and Electrical Insulation* **16** 281–288.

[8] COX, D. R. and OAKES, D. (1984). *Analysis of Survival Data.* Chapman & Hall, New York.. MR0751780

[9] DAVISON, A. C. and HINKLEY, D. V. (1997). *Bootstrap Methods and Their Application.* Cambridge University Press, Cambridge. MR1478673

[10] DE CARVALHO, M. (2011). Confidence intervals for the minimum of a function using extreme value statistics. *International Journal of Mathematical Modelling & Numerical Optimisation* **2** 288–296.

[11] DE HAAN, L. (1981). Estimation of the minimum of a function using order statistics. *Journal of the American Statistical Association* **76** 467–469. MR0624350

[12] EFRON, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics* **7** 1–26. MR0515681

[13] EFRON, B., HASTIE, T., JOHNSTONE, L., and TIBSHIRANI, R. (2004). Least angle regression (with discussion). *Annals of Statistics* **32** 407–451. MR2060166

[14] ETHIER, S. N. (1982). Testing for favorable numbers on a roulette wheel. *Journal of the American Statistical Association* **77** 660–665. MR0675895

[15] FAN, J., GUO, S., and HAO, N. (2012). Variance estimation using refitted cross-validation in ultrahigh dimensional regression. *Journal of the Royal Statistical Society, Ser. B* **74** 37–65. MR2885839

[16] FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* **96** 1348–1360. MR1946581

[17] FAN, J., ZHANG, C., and ZHANG J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *Annals of statistics* **29** 153–193. MR1833962

[18] FANG, K. T., HICKERNELL, F. J., and WINKER, P. (1996). Some global optimization algorithms in statistics, in *Lecture Notes in Operations Research*, eds. by Du, D. Z., Zhang, X. S. and Cheng, K. World Publishing Corporation, 14–24.

[19] FANG, K. T., LIN, D. K. J., WINKER, P., and ZHANG, Y. (2000). Uniform design: theory and application. *Technometrics* **42** 237–248. MR1801031

[20] FISHER, R. A. (1959). *Statistical Methods and Scientific Inference*, 2nd ed., revised. Hafner Publishing Company, New York. MR0131909

[21] FRANK, L. E. and HEISER, W. J. (2008). Feature selection in feature network models: Finding predictive subsets of features with the positive Lasso. *British Journal of Mathematical and Statistical Psychology* **61** 1–27. MR2522109

[22] GELFAND, A. E., GLAZ, J., KUO, L., and LEE, T. M. (1992). Inference for the maximum cell probability under multinomial sampling. *Naval Research Logistics* **39** 97–114.

[23] GHOSH, J. K., DELAMPADY, M., and SAMANTA, T. (2007). *An Introduction to Bayesian Analysis: Theory and Methods.* Springer, New York. MR2247439

[24] GLAZ, J. and SISON, C. P. (1999). Simultaneous confidence intervals for multinomial proportions. *Journal of Statistical Planning and Inference* **82** 251–262. MR1736447

[25] HALL, P. (1992). *The bootstrap and Edgeworth Expansion.* Springer, New York. MR1145237

[26] HANNIG, J., IYER, H., and PATTERSON, P. (2006). Fiducial feneralized confidence intervals. *Journal of the American Statistical Association* **101** 254–269. MR2268043

[27] HÄRDLE, W. and LUCKHAUS, S. (1984). Uniform consistency of a class of regression function estimators. *Annals of Statistics* **12** 612–623. MR0740915

[28] HART, J. D. (1997). *Nonparametric Smoothing and Lack-of-Fit Tests.* Springer, New York. MR1461272

[29] HOMEM-DE-MELLO T. (2008). On rates of convergence for stochastic optimization problems under non-independent and identically distributed sampling. *SIAM Journal on Optimization* **19** 524–551. MR2425028

[30] JOHNSON, M. E., MOORE, L. M., and YLVISAKER, D. (1990). Minimax and maximin distance designs. *Journal of Statistical Planning and Inference* **26** 131–148. MR1079258

[31] KUSHNER, H. J. and YIN, G. G. (1997). *Stochastic Approximation Algorithms and Applications.* Springer, New York. MR1453116

[32] LEHMANN, E. L. and ROMANO, J. P. (2006). *Testing Statistical Hypotheses*, 3rd., Springer, Science & Business Media. MR2135927

[33] LOCKHART, R. A. and STEPHENS, M. A. (1994). Estimation and tests of fit for the three-parameter Weibull distribution. *Journal of the Royal Statistical Society, Ser. B* **56** 491–500. MR1278222

[34] MARTIN, R. (2012). Plausibility functions and exact frequentist inference. *Journal of the American Statistical Association* **110** 1552–1561. MR3449054

[35] MCKAY, M. D., BECKMAN, R. J., and CONOVER, W. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics* **21** 239–245. MR0533252

[36] MENG, X-L. (1994). Posterior predictive *p*-values. *Annals of Statistics* **22** 1142–1160. MR1311969

[37] MURTHY, D. N. P., XIE, M., and JIANG, R. (2004). *Weibull Models.* John Wiley & Sons, Hoboken, New Jersey. MR2013269

[38] PARK, J. S. (2001). Optimal Latin-hypercube designs for computer experiments. *Journal of Statistical Planning Inference* **39** 15–111. MR1266995

[39] PATIL, G. P. and TAILLIE, C. (1979). An overview of diversity. *Ecological Diversity in Theory and Practice*, ed. by Grassle, J. F. et al. International Co-operative Publishing House. Fairland, MD.

[40] POLITIS, D. N., ROMANO, J. P., and WOLF, M. (1999). *Subsampling.* Springer, New York. MR1707286

[41] SHAO, J. and TU, D. (1995). *The Jackknife and Bootstrap.* Springer, New York. MR1351010

[42] SHAPIRO, A. (2003). Monte Carlo sampling methods, in Stochastic Programming. *Handbook in OR & MS*, Vol. 10, ed. by Ruszczyński, A. and Shapiro, A., North-Holland, Amsterdam. MR2051793

[43] SILVAPULLE, M. J. and SEN, P. K. (2011). *Constrained Statistical Inference: Order, Inequality, and Shape Constraints.* John Wiley & Sons, Hoboken, New Jersey. MR2099529

[44] TEIMOURI, M., HOSEINI, S. M., and NADARAJAH, S. (2013). Comparison of estimation methods for the Weibull distribution. *Statistics: A Journal of Theoretical and Applied Statistics* **47** 93–109. MR3023017

[45] TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Ser. B* **58** 267–288. MR1379242

[46] WEERAHANDI, S. (1995). *Exact Statistical Methods for Data Analysis.* Springer, New York. MR1316663

[47] WU, L., YANG, Y., and LIU, H. (2014). Nonnegative-lasso and application in index tracking. *Computational Statistics & Data Analysis* **70** 116–126 MR3125482

[48] XIE, M., SINGH, K., and STRAWDERMAN, W. E. (2011). Confidence distributions and a unifying framework for meta-analysis. *Journal of the American Statistical Association* **106** 320–333. MR2816724

[49] XIONG, S. and LI, G. (2009). Inference for ordered parameters in multinomial distributions. *Science in China, Ser. A: Mathematics* **52** 526–538. MR2491771

[50] XIONG, S. and LI, G. (2010). Dispersion comparisons of two probability vectors under multinomial sampling. *Acta Mathematica Scientia* **30** 907–918. MR2675796

[51] ZHANG. C-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Annals of Statistics* **38** 894–942. MR2604701