Journal of Multivariate Analysis 132 (2014) 171–182

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

# Smoothed and iterated bootstrap confidence regions for parameter vectors

Santu Ghosh *, Alan M. Polansky

*Wayne State University, USA*
*Northern Illinois University, USA*

ARTICLE INFO

*Article history:*
Received 26 December 2013
Available online 27 August 2014

*AMS subject classifications:*
primary 62G15
secondary 62G09

*Keywords:*
Bandwidth matrix
Bootstrap percentile method
Bootstrap percentile-$t$ method
Iterated bootstrap method
Edgeworth expansion
Smooth function model

ABSTRACT

The construction of confidence regions for parameter vectors is a difficult problem in the nonparametric setting, particularly when the sample size is not large. We focus on bootstrap ellipsoidal confidence regions. The bootstrap has shown promise in solving this problem, but empirical evidence often indicates that the bootstrap percentile method has difficulty in maintaining the correct coverage probability, while the bootstrap percentile-$t$ method may be unstable, often resulting in very large confidence regions. This paper considers the smoothed and iterated bootstrap methods to construct the bootstrap percentile method ellipsoidal confidence region. The smoothed bootstrap method is based on a multivariate kernel density estimator. An optimal bandwidth matrix is established for the smoothed bootstrap procedure that reduces the coverage error of the bootstrap percentile method. We also provide an analytical adjustment to the nominal level to reduce the computational cost of the iterated bootstrap method. Simulations demonstrate that the methods can be successfully applied in practice.



## 1. Introduction

The construction of bootstrap confidence intervals has been studied extensively over the past few decades. Early criticism of the bootstrap percentile method (Efron, [4]) led to several improvements of the methodology, including the bias corrected method (Efron, [5]), the bias-corrected and accelerated method (Efron, [7]), and the studentized method (Efron, [6]). Methods based on pre-pivoting, the iterated bootstrap, and calibration were developed by Beran [2], Hall [9], and Loh [14]. Hall [10] provided a systematic method for comparing confidence intervals based on Edgeworth expansion theory. Implementation of the smoothed bootstrap with the specific purpose of improving the coverage properties of confidence intervals has been discussed by Guerra, Polansky and Schucany [8], Polansky [16], and Polansky and Schucany [17]. However, multivariate confidence regions have received limited consideration and it is difficult to extend most of the existing univariate procedures directly to the multivariate case.

Let $X_1, X_2, \ldots, X_n$ be a set of independent and identically distributed $p$-dimensional random vectors following a distribution $F$. Let $\theta = t(F)$ be a $v$-dimensional parameter vector, $\hat{\theta}_n$ is a plug-in estimator of $\theta$, and $\hat{\Omega}_n$ is a consistent estimator of the asymptotic covariance matrix $\Omega$ of $n^{1/2}\hat{\theta}_n$. Assume that $\Omega$ is non-singular. Then a $100\alpha\%$ confidence region for $\theta$ has the form

$$\mathcal{R} = \{\hat{\theta}_n - n^{-1/2}\hat{\Omega}_n^{1/2} r : r \in \mathcal{R}_\alpha\},$$

* Corresponding author at: Wayne State University, USA.
*E-mail address:* santughosh001@gmail.com (S. Ghosh).

http://dx.doi.org/10.1016/j.jmva.2014.08.003

where $\mathcal{R}_\alpha \subset \mathbb{R}^v$ is any region such that $P[\sqrt{n}\hat{\Omega}_n^{-1/2}(\hat{\theta}_n - \theta) \in \mathcal{R}_\alpha] = \alpha$. The shape of the region $\mathcal{R}$ depends on the shape of the region $\mathcal{R}_\alpha$. In this paper we concentrate on ellipsoidal confidence regions, which are generalizations of univariate symmetric confidence intervals. In particular, if $\mathcal{R}_\alpha$ is a $v$-variate sphere centered at the origin, then $\mathcal{R}$ becomes an ellipsoidal confidence region. In practice the bootstrap is often used to estimate $\mathcal{R}_\alpha$.

A simpler method for computing an ellipsoidal confidence region for $\theta$ is based on extending the bootstrap percentile method of Efron [4] to the multivariate case. Let $\mathcal{R}_{BP}$ be a bootstrap percentile method ellipsoidal confidence region for $\theta$. For a given nominal level $\alpha$, we will prove that

$$P(\theta \in \mathcal{R}_{BP}) = \alpha + n^{-1}Q(\chi^2_{v,\alpha}) + O(n^{-2}), \tag{1.1}$$

where $Q(\chi^2_{v,\alpha})$ is a polynomial in $\chi^2_{v,\alpha}$. The coefficients of $Q(\chi^2_{v,\alpha})$ are functions of population moments and $\chi^2_{v,\alpha}$ is the $\alpha$-th quantile of a chi-square distribution with $v$ degrees of freedom. Eq. (1.1) shows that $\mathcal{R}_{BP}$ is second-order accurate. Our empirical studies show that this method is stable, but has poor coverage properties. An alternative method for constructing an ellipsoidal confidence region for $\theta$ is the bootstrap percentile-$t$ method. From an asymptotic viewpoint, the bootstrap percentile-$t$ method is fourth-order accurate, see Hall [11, Section 4.2]. Our empirical studies show that while the bootstrap percentile-$t$ method has acceptable coverage probabilities, it can be unstable and can produce large ellipsoidal confidence regions when the sample size is small.

The natural idea is to improve the coverage probability of $\mathcal{R}_{BP}$. In the univariate setup, the smoothed and iterated bootstrap methods have potential applications in the construction of confidence intervals. Both of these methods are easily implementable as practical procedures for routine use. To our knowledge, so far the use of the smoothed and iterated bootstrap methods have not been explored in the case of multivariate regions. To improve the coverage probability of $\mathcal{R}_{BP}$, we consider a multivariate version of the smoothed and iterated bootstrap methods. However, the performance of the smoothed bootstrap heavily depends on the choice of the bandwidth matrix and the latter method is computationally expensive, specifically in the multivariate case.

In this paper our contributions are: (i) we show that a bandwidth matrix of order $O(1/n)$ exists that eliminates the $n^{-1}$-term from Eq. (1.1). As a result, the bootstrap percentile method based on the smoothed bootstrap with this bandwidth matrix becomes fourth-order accurate and (ii) an analytical correction is provided to the nominal level in order to avoid the double bootstrap for constructing the iterated bootstrap percentile method ellipsoidal confidence region. We also show that the resulting region reduces the order of the coverage error of $\mathcal{R}_{BP}$ to $O(n^{-2})$.

The remainder of the paper is organized as follows. Section 2 introduces the smoothed and iterated bootstrap methods for the mean vector. In Section 3, we extend these methods for a multivariate smooth function of a mean vector. Simulation results are reported in Section 4. Section 5 concludes and Appendix contains some technical details.

## 2. Bootstrap confidence regions for a mean vector

Let $\theta = E_F(X_n)$ be a $p$-dimensional mean vector of $F$ and assume that the covariance matrix $\Sigma$, of $F$, is positive definite and unknown. We are interested in constructing an ellipsoidal confidence region for $\theta$. Let

$$\hat{\theta}_n = n^{-1}\sum_{i=1}^{n} X_i, \quad \text{and} \quad \hat{\Sigma}_n = n^{-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)(X_i - \bar{X}_n)'.$$

To facilitate our discussion of the bootstrap percentile method ellipsoidal confidence region, let $X_1^*, \ldots, X_n^*$ be a random sample from the empirical distribution $\hat{F}_n$, based on $X_1, \ldots, X_n$. Let

$$\hat{\theta}_n^* = n^{-1}\sum_{i=1}^{n} X_i^*. \quad \text{and} \quad \hat{\Sigma}_n^* = n^{-1}\sum_{i=1}^{n}(X_i^* - \bar{X}_n^*)(X_i^* - \bar{X}_n^*)'.$$

A bootstrap percentile method ellipsoidal confidence region for $\theta$ with approximate coverage probability $\alpha$ is

$$\mathcal{R}_{BP} = \{\hat{\theta}_n - n^{-1/2}\hat{\Sigma}_n^{1/2}s : s \in \mathcal{S}_{BP,\alpha}\},$$

where $\mathcal{S}_{BP,\alpha}$ denotes a $p$-variate sphere centered at the origin such that $P^*(S^* \in \mathcal{S}_{BP,\alpha}) = \alpha$ and $S^* = \sqrt{n}\hat{\Sigma}_n^{-1/2}(\hat{\theta}_n^* - \hat{\theta}_n)$. $P^*$ denotes the probability measure conditional on $X_1, \ldots, X_n$.

An alternative method is the *bootstrap percentile-t ellipsoidal confidence region* for $\theta$ with approximate coverage probability $\alpha$, given by

$$\mathcal{R}_{BT} = \{\hat{\theta}_n - n^{-1/2}\hat{\Sigma}_n^{1/2}s : s \in \mathcal{S}_{BT,\alpha}\},$$

where $\mathcal{S}_{BT,\alpha}$ denotes a $p$-variate sphere centered at the origin such that $P^*(U^* \in \mathcal{S}_{BT,\alpha}) = \alpha$ and $U^* = \sqrt{n}\hat{\Sigma}_n^{*-1/2}(\hat{\theta}_n^* - \hat{\theta}_n)$. $\mathcal{R}_{BT}$ can be unstable if there is a significant conditional probability under $\hat{F}_n$ that $\hat{\Sigma}_n^*$ is nearly singular. We begin with the asymptotic expansion for the coverage probability of $\mathcal{R}_{BP}$. The following assumptions are made throughout this section:

1. The distribution $G$ of $Y = \begin{bmatrix}\text{vec}(X)\\ \text{vech}(XX')\end{bmatrix}$ satisfies the multivariate version of the Cramér continuity condition. The condition holds provided $G$ has a non-degenerate absolutely continuous component. See Hall [11, pp. 66–67].
2. Assume all moments of order 6 of $Y$ are finite. That is $E(\|Y\|^6) < \infty$.