Opuscula Math. 41, no. 5 (2021), 725–740
https://doi.org/10.7494/OpMath.2021.41.5.725 Opuscula Mathematica

NONPARAMETRIC BOOTSTRAP CONFIDENCE BANDS FOR UNFOLDING SPHERE SIZE DISTRIBUTIONS

Jakub Wojdyła

Communicated by Mirosław Pawlak

Abstract. The stereological inverse problem of unfolding the distribution of spheres radii from measured planar sections radii, known as the Wicksell's corpuscle problem, is considered. The construction of uniform confidence bands based on the smoothed bootstrap in the Wicksell's problem is presented. Theoretical results on the consistency of the proposed bootstrap procedure are given, where the consistency of the bands means that the coverage probability converges to the nominal level. The finite-sample performance of the proposed method is studied via Monte Carlo simulations and compared with the asymptotic (non-bootstrap) solution described in literature.

Keywords: bootstrap, confidence bands, inverse problem, nonparametric density estimation, Wicksell's problem.

Mathematics Subject Classification: 45Q05, 62G05, 62G15, 62G20.

1. INTRODUCTION

This article concerns the construction of bootstrap simultaneous confidence intervals (confidence bands) in a certain statistical problem, which, from mathematical point of view, is an inverse problem, and was originally motivated by medical applications.

The Wicksell's corpuscle problem ([28]) is an example of a classical problem of stereology. We deal with it when we model an experiment in which objects in the shape of balls of random radii are randomly placed in some three-dimensional opaque medium and we want to know the distribution of the spheres radii. Assume that the lengths of the radii are independent and have the same distribution. We cut the medium with a random plane and measure the observed circles radii. The goal is to unfold the distribution of spheres radii from the measured planar sections radii. Examples of real life applications of Wicksell's corpuscle problem are seen in, e.g., geology (where mineral grains in a rock are considered to be the spheres), metallurgy (where we consider graphite grains as balls of varying radii), or medicine (where cancer cells in the tissue can be treated as spheres). Medicine is on this list also because of historical reasons, as it was primary motivation of Wicksell's work. In geology, and similarly in metallurgy, we would like to know the size of the spheres because it affects the properties of the material, like fragility. As for medicine the purpose is obvious. The problem arises in numerous other contexts (biology, engineering, astronomy, etc.), see, e.g., [7,8] and the references given there. In [7], the Wicksell's problem formulation using marked point processes formalism can be found.

Throughout this article, as in, e.g., [1, 17, 20], we consider squared radii of both the unobserved spheres and the observed circular sections. It is more convenient mathematically and sometimes it is also easier to measure the squared radius than the radius itself (see [20]). It is clear, however, that the results obtained for squared radii can be easily transformed back to the original problem.

Let f and g denote, respectively, the density of the squared spheres radii and the density of the observable squared circles radii. The relation between them is given by the equation (see, e.g., [18, Sec. 4.1] or [20])

$$g(z) = \frac{1}{2m} \int_z^\infty (x - z)^{-1/2} f(x)\, dx, \quad z \geq 0, \tag{1.1}$$

where $m = 2\int_0^\infty x^2 f(x^2)\, dx$ is the mean sphere radius. The solution, when it exists, is given by

$$f(x) = \frac{-2m}{\pi} \frac{d}{dx} \int_x^\infty (z - x)^{-1/2} g(z)\, dz, \quad x \geq 0. \tag{1.2}$$

We assume that the density f, and hence also g, has a bounded support $[0, 1]$. In standard L_2-settings (with temporarily fixed m), the operator defined by equation (1.2) is unbounded, which makes the Wicksell's problem ill-posed in the Hadamard sense, and some sort of regularization is required.

Using confidence bands is the most informative way of quantifying the accuracy of estimators in problems of function estimation. In particular, they graphically illustrate how the accuracy of the estimation varies with x. A lot of work has been devoted to the construction of confidence bands in direct problems of function estimation, starting with the pioneering work of Bickel and Rosenblatt in 1973 ([2]), who constructed confidence bands for density estimated from an i.i.d. sample. For reviews of papers on this topic, see, e.g., [16, Chapter 5.1.3], or [3, 4, 24]. To the best of our knowledge, article [4] published in 2007 was the first work that formally studied confidence bands in inverse problems. Since then, several related works have been published, see, e.g., [3, 5, 6, 10, 12, 21–24, 30]. Despite their practical importance, confidence bands in stereological inverse problems haven't been constructed until only recently (see [10, 30]). In particular, in [30] such bands were produced in the Wicksell's problem, considered in this paper, however, without bootstrapping.

In [30], similarly to [4], a Bickel-Rosenblatt type limit theorem was proven, and the asymptotic confidence bands were constructed based on it. Moreover, to improve the finite sample coverage properties of the confidence bands, the authors of [4] used

Efron's original bootstrap algorithm to construct percentile bootstrap confidence bands for the density of interest. The fact that a bootstrap approach proved to be successful in a similar density deconvolution problem was an additional motivation to apply this methodology in the Wicksell's problem.

From the theoretical point of view, it is possible to directly adapt the construction of the bootstrap bands from [4] for the case of the Wicksell's problem. However, extensive numerical experiments, not reported here, showed that the standard bootstrap doesn't perform better than the asymptotic approach, at least in the considered cases. The remedy to that was to use the so-called smoothed bootstrap, which is an extension of standard bootstrap procedure, where instead of drawing samples with replacement from the empirical distribution, they are drawn from a kernel density estimate of the distribution. The authors of, e.g., [11,13,14,19,25] wrote about the potential advantage of this version of the bootstrap over the standard version. A theoretical introduction to the topic can be found in, e.g., [15]. The resulting smoothed bootstrap confidence bands in the Wicksell's problem are also asymptotically consistent, and the proof of their consistency is presented in the next section.

The rest of the paper is organized as follows. Section 2 presents the main theoretical results of this work, our methodology of constructing bootstrap confidence bands for the density f. In Section 3, we present results of simulation studies conducted to verify the finite sample performance of the proposed bootstrap confidence bands, in particular to compare bootstrap and asymptotic (non-bootstrap) approaches.

2. CONFIDENCE BANDS BASED ON THE BOOTSTRAP

The centre of the confidence bands is determined by an appropriate version of kernel-type estimator proposed in [27]. Let $Z_1, \ldots, Z_n$ denote the observed squared radii of circular profiles, and let G denote a distribution function of $Z_1, \ldots, Z_n$. Consider a density estimator of f given by the formula

$$f_n(x) = \frac{-2m}{nh^{3/2}\pi} \sum_{i=1}^{n} K\left(\frac{x - Z_i}{h}\right), \quad x \geq 0, \tag{2.1}$$

with

$$K(x) = \int_0^\infty y^{-1/2} K_0'(y + x)\, dy, \quad x \in \mathbb{R}, \tag{2.2}$$

where K_0 is a kernel function satisfying assumption 1 imposed later in this work, and h is a bandwidth satisfying assumption 2. This is the form one gets by inserting a standard kernel-type estimator of the density g into equation (1.2) describing the relation between f and g (cf. also [30]). The estimator defined this way depends on the unknown parameter m, which, in our theoretical results and simulations, will be replaced with an appropriate estimator $\hat{m}$.

2.1. BOOTSTRAPPING

Our method builds on the smoothed bootstrap. Let $Z_1^*, \ldots, Z_n^*$ denotes a random sample simulated from $\bar{G}_{n,\eta}$, i.e. the kernel-smoothed version of the empirical distribution function of $Z_1, \ldots, Z_n$ with density given by the kernel density estimator

$$\bar{g}_{n,\eta}(x) = \frac{1}{n\eta} \sum_{i=1}^{n} \bar{K}_0 \left(\frac{x - Z_i}{\eta} \right),$$

with a kernel function $\bar{K}_0$ and a bandwidth η. Let P^* denotes the conditional probability, given $Z_1, \ldots, Z_n$, and let f_n^* be the bootstrap version of the estimator f_n defined by (2.1):

$$f_n^*(x) = \frac{-2m}{nh^{3/2}\pi} \sum_{i=1}^{n} K \left(\frac{x - Z_i^*}{h} \right), \quad x \geq 0.$$

Moreover, define the process

$$Y_n^*(t) = -\frac{n^{1/2}h\pi}{2m\tilde{g}_n(t)^{1/2}} \left[f_n^*(t) - f_n(t) \right], \quad t \in [a,b],$$

as the bootstrap approximation of the process

$$Y_n(t) = -\frac{n^{1/2}h\pi}{2mg(t)^{1/2}} \left[f_n(t) - E\left\{ f_n(t) \right\} \right], \quad t \in [a,b],$$

for which the limiting distribution of the supremum was investigated in [30]. Here, $\tilde{g}_n$ is an appropriate estimator of the density g. It will be proven that the process Y_n^* has the same limiting distribution as the process Y_n.

2.2. ASSUMPTIONS

The theory is developed under the following assumptions. Confidence bands will be constructed on the interval $[a,b]$ which is a compact subset of $[0,1]$, with $a > 0$, $b < 1$. Denote by $\|\cdot\|$ the sup-norm on the interval $[a,b]$. Throughout this article, we define the order of a kernel as the order of its first non-zero moment.

We require the kernel K_0 to belong to the class defined by assumptions 1(a), 1(b), and the kernel $\bar{K}_0$ to satisfy assumption 1(c).

Assumption 1.

(a) For some integer $k \geq 1$, K_0 is a kernel of order at least k, supported and differentiable on $[-1,1]$.

(b) The kernel K defined in (2.2) is differentiable, integrable, square integrable, and satisfies

$$K(x)|x|^{1/2}[\log\log|x|]^{1/2} \to 0, \quad \text{when} \quad |x| \to \infty.$$

Moreover, K' is square integrable and for some $\alpha > 0$ satisfies

$$\int |K'(x)||x|^{1/2+\alpha}\, dx < \infty.$$

(c) The kernel $\bar{K}_0$ is a density function such that

$$\int x\bar{K}_0(x)\ dx = 0 \quad \text{and} \quad \int x^2\bar{K}_0(x)\ dx < \infty.$$

The bandwidths h (used in the main part of the construction) and η (from the bootstrap procedure) are assumed to tend to zero, as $n \to \infty$, in such a way that assumption 2 holds.

Assumption 2.

(a) $n^{-1/2+\delta}\log(1/h) = O(1)$,
(b) $n^{\delta}h\log(1/h)^{1/2} = O(1)$ for some $\delta > 0$,
(c) $n^{1/2}h^{k+1}\log(1/h)^{1/2} \to 0$ for k from the assumption 1(a),
(d) $nh^{2+\theta}/\log(1/h) \to \infty$ for some $\theta > 0$,
(e) $(n/\log\log n)^{1/2}\eta^2 \to 0$.

Furthermore, when constructing the bands, one needs to estimate the unknown density g of the observations and the unknown mean m. We assume that this is done using estimators satisfying assumption 3.

Assumption 3.

(a) The estimator $\tilde{g}_n$ of the unknown density g of the observations satisfies

$$\|\tilde{g}_n - g\| = o\left(1/\log(1/h)\right) \quad \text{almost surely},$$

where h is the bandwidth chosen for the construction of the estimators f_n and f_n^* of f.

(b) The estimator $\hat{m}$ of the unknown mean m is such that, for all $\varepsilon > 0$,

$$\hat{m} - m = O_p(n^{-1/2+\varepsilon}).$$

Finally, we make the following assumptions about the density of interest f and the density g (related to f by equation (1.1)).

Assumption 4.

(a) The density g is bounded away from zero on $[a, b]$ and $g^{1/2}$ is differentiable with bounded derivative on $[0, 1]$.
(b) For k as in assumption 1(a), f is $(k-1)$-times continuously differentiable in $[0, 1]$ and there exists bounded $f^{(k)}$ in $(0, 1)$.

Assumption 1(a) coincides exactly with assumption (1a) from [30], while 1(b) with (1b). An example of a kernel function that satisfies assumptions 1(a) and 1(b), with $k = 2$, is the biweight kernel

$$K_0(x) = (15/16)(1 - x^2)^2 I_{[-1,1]}(x).$$

Assumption 1(c) relates to the regularity of the kernel function used in the bootstrap procedure and, together with assumptions 2(e) and 4(a), guarantees the appropriate convergence rate of $\bar{G}_{n,\eta}$ to the true distribution function G ([29]). Assumptions 2(a), 2(b), 2(c), about the convergence rate of the smoothing parameter h, are among the assumptions from Corollary 1 in [30]. The assumption 2(d) is a slightly stronger version of the corresponding assumption from Corollary 1 in [30]. Assumption 2(e) concerns the convergence rate of the smoothing parameter used to generate the bootstrap sample and, as it was mentioned before, guarantees the appropriate properties of the bootstrap distribution function. Assumption 3(a) is a stronger version of the condition (2.4) from [30] and is satisfied for, e.g., kernel density estimators with the smoothing parameter h, if the kernel function is of bounded variation and has compact support, and the estimated density g is differentiable (see [26]). An example of an estimator with the property 3(b) is given in [30]:

$$\hat{m} = \frac{n\pi}{2}\left(\sum_{i=1}^{n} Z_i^{-1/2}\right)^{-1}.$$

Assumptions 4(a) and 4(b) are exactly the same as assumptions (2a) and (2b) in [30].

2.3. MAIN RESULTS

In the following theorem we establish the bootstrap analogue of Theorem 1 in [30]. The idea of the proof of the consistency of the proposed bootstrap bands, based on so-called strong approximation of the appropriate empirical process, was taken from [4]. The reasoning from [4] required, however, technical modifications, mainly due to a different form of the estimator in the Wicksell's problem and due to the use of the smoothed version of bootstrap (instead of the standard bootstrap, as in [4]).

Theorem 2.1. *Under assumptions 1(b), 2(d), and 4(a), for each $x \in \mathbb{R}$,*

$$P^*\left([2\log(1/h)]^{1/2}\left[\|Y_n^*\|/C_{K,1}^{1/2} - d_n\right] < x\right) \to \exp\{-2\exp(-x)\}$$

almost surely, where

$$d_n = [2\log(1/h)]^{1/2} + \frac{\log\{C_{K,2}^{1/2}/(2\pi)\}}{[2\log(1/h)]^{1/2}},$$

and

$$C_{K,1} = \int K(x)^2\,dx, \quad C_{K,2} = \frac{b-a}{C_{K,1}}\int K'(x)^2\,dx.$$

Proof. Let $U_1^*, U_2^*, \ldots$ be i.i.d. random variables uniformly distributed on $[0,1]$, and let $\alpha_n^*(t) = n^{1/2}\left[U_n(t) - t\right]$ be the corresponding empirical process, where U_n is the empirical distribution function based on $U_1^*, \ldots, U_n^*$. The Komlós–Major–Tusnády approximation (see, e.g., Theorem 4.4.1 in [9]) gives the existence of a sequence of Brownian bridges B_n^* such that

$$\sup_{t\in[0,1]} |\alpha_n^*(t) - B_n^*(t)| = O_{p^*}(n^{-1/2}\log n), \quad n\to\infty, \quad \text{almost surely}$$

(for almost all sequences of the observations $Z_1, Z_2, \ldots$). Assume that both the variables U_i^* and the processes B_n^* are independent of the observations Z_i. Furthermore, without loss of generality, one can assume that $Z_i^* = \bar{G}_{n,\eta}^{-1}(U_i^*)$, $i = 1, \ldots, n$. Let $\bar{G}_{n,\eta}^*$ be the empirical distribution function based on $Z_1^*, \ldots, Z_n^*$ and let

$$\alpha_n^{\bar{G}_{n,\eta}}(t) = n^{1/2}\left[\bar{G}_{n,\eta}^*(t) - \bar{G}_{n,\eta}(t)\right]$$

be the bootstrap empirical process. Then, by applying the argument from the proof of Theorem 2 in [4], and the fact that the distribution function $\bar{G}_{n,\eta}$ has the so-called Chung–Smirnov property, i.e.

$$\sup_{x\in\mathbb{R}} |\bar{G}_{n,\eta}(x) - G(x)| = O([n^{-1}\log(\log^+ n)]^{1/2}) \quad \text{almost surely}$$

(see Theorem 3.2 in [29]), one can deduce that, with any $0 < \beta < 1/2$,

$$\sup_{t\in\mathbb{R}} |\alpha_n^{\bar{G}_{n,\eta}}(t) - B_n^*\{G(t)\}| = O_{p^*}(n^{-\beta/2}\log n), \quad n \to \infty, \quad \text{almost surely.} \tag{2.3}$$

Let

$$Y_{n,0}(t) = \frac{h^{-1/2}}{g(t)^{1/2}} \int_0^1 K\left(\frac{t-x}{h}\right) dB_n\{G(x)\}, \quad t \in [a,b],$$

be the process defined in the proof of Theorem 1 in [30] (where B_n is a Brownian bridge), for which, for each $x \in \mathbb{R}$,

$$P\left([2\log(1/h)]^{1/2}\left[\|Y_{n,0}\|/C_{K,1}^{1/2} - d_n\right] \le x\right) \to \exp\{-2\exp(-x)\} \tag{2.4}$$

(note that the assumptions of Theorem 1 in [30] are satisfied because of assumptions 1(b), 4(a) and assumption 2(d), which implies that $nh/(\log n)^3 \to \infty$). We introduce two processes that approximate Y_n^*:

$$Y_{n,0}^*(t) = \frac{h^{-3/2}}{g(t)^{1/2}} \int_0^1 K'\left(\frac{t-x}{h}\right) B_n^*\{G(x)\}\, dx, \quad t \in [a,b],$$

$$Y_{n,4}^*(t) = -\frac{n^{1/2}h\pi}{2mg(t)^{1/2}}[f_n^*(t) - f_n(t)], \quad t \in [a,b],$$

which are the bootstrap analogues of the processes $Y_{n,0}$ (see above) and $Y_{n,4}$ appearing in the construction of asymptotic confidence bands for f in [30].

Taking the difference of these processes, one has

$$\left|Y_{n,0}^*(t) - Y_{n,4}^*(t)\right| \le \frac{h^{-1/2}}{g(t)^{1/2}} \sup_{x\in\mathbb{R}} \left|\alpha_n^{\bar{G}_{n,\eta}}(x) - B_n^*\{G(x)\}\right| \int |K'(u)|\, du$$

almost surely, and, hence,

$$\|Y_{n,0}^* - Y_{n,4}^*\| = O_{p^*}(n^{-\beta/2}h^{-1/2}\log n) = o_{p^*}\{\log(1/h)^{-1/2}\} \quad \text{almost surely,} \tag{2.5}$$

by an application of assumption 4(a), equation (2.3), assumption 1(b), and assumption 2(d).

Because

$$\mathcal{L}^* \left([2\log(1/h)]^{1/2} \left[\|Y^*_{n,0}\|/C^{1/2}_{K,1} - d_n \right] \right)$$
$$= \mathcal{L} \left([2\log(1/h)]^{1/2} \left[\|Y_{n,0}\|/C^{1/2}_{K,1} - d_n \right] \right) \quad \text{almost surely,}$$

it follows from equation (2.4) that

$$P^* \left([2\log(1/h)]^{1/2} \left[\|Y^*_{n,0}\|/C^{1/2}_{K,1} - d_n \right] \leq x \right) \to \exp\{-2\exp(-x)\}$$

almost surely. Consequently, equation (2.5) implies that

$$P^* \left([2\log(1/h)]^{1/2} \left[\|Y^*_{n,4}\|/C^{1/2}_{K,1} - d_n \right] \leq x \right) \to \exp\{-2\exp(-x)\} \tag{2.6}$$

almost surely.

Furthermore,

$$Y^*_n(t) - Y^*_{n,4}(t) = \frac{g(t) - \tilde{g}_n(t)}{\tilde{g}_n(t)^{1/2}[g(t)^{1/2} + \tilde{g}_n(t)^{1/2}]} Y^*_{n,4}(t)$$

and, hence,

$$\|Y^*_n - Y^*_{n,4}\| = o_{p^*}\{\log(1/h)^{-1/2}\} \quad \text{almost surely} \tag{2.7}$$

because of assumption 3(a) and of $\|Y^*_{n,4}\| = O_{p^*}\{\log(1/h)^{1/2}\}$ almost surely, which follows from equation (2.6).

Theorem 2.1 now follows from equations (2.6) and (2.7). □

The above result together with Corollary 1 from [30] show that, under appropriate conditions, $\|Y^*_n\|$ and $\|Y_{n,6}\|$ have the same limiting distribution, where

$$Y_{n,6}(t) = -\frac{n^{1/2}h\pi}{2\hat{m}\tilde{g}_n(t)^{1/2}}[\hat{f}_n(t) - f(t)], \quad t \in [a, b],$$

and $\hat{f}_n$ denotes the estimator obtained from f_n by replacing m in the definition (2.1) with the estimator $\hat{m}$. From this, one directly obtains the following bootstrap confidence bands.

Corollary 2.2. *Under assumptions 1-4,*

$$\left\{ \left[\hat{f}_n(t) - \frac{2\hat{m}\tilde{g}_n(t)^{1/2}q^*_{1-\alpha}}{n^{1/2}h\pi}, \hat{f}_n(t) + \frac{2\hat{m}\tilde{g}_n(t)^{1/2}q^*_{1-\alpha}}{n^{1/2}h\pi} \right], t \in [a, b] \right\} \tag{2.8}$$

is a consistent $(1-\alpha) \times 100\%$ *asymptotic bootstrap confidence band for* f *over* $[a, b]$, *where* $q^*_{1-\alpha}$ *denotes the* $(1-\alpha)$*-quantile of the bootstrap distribution of* $\|Y^*_n\|$.

Proof. Let $H(x) = \exp\{-2\exp(-x)\}$ be the distribution function of the asymptotic distribution from Theorem 2.1, and let H_n and H_n^* denote the distribution functions of

$$[2\log(1/h)]^{1/2}[\|Y_{n,6}\|/C_{K,1}^{1/2} - d_n] \quad \text{and} \quad [2\log(1/h)]^{1/2}[\|Y_n^*\|/C_{K,1}^{1/2} - d_n],$$

respectively. Furthermore, denote with q and q_n^* the quantile functions corresponding to, respectively, H and H_n^*. Then, for each $x \in (0,1)$,

$$q_n^*(x) \to q(x) \quad \text{almost surely,} \tag{2.9}$$

since, for each $x \in \mathbb{R}$, $H_n^*(x) \to H(x)$ almost surely, and

$$\sup_{x\in\mathbb{R}} |H_n(x) - H(x)| \to 0. \tag{2.10}$$

From equations (2.9) and (2.10) one obtains, for each $x \in (0,1)$,

$$\begin{aligned} |H_n\{q_n^*(x)\} - H\{q(x)\}| &\le |H_n\{q_n^*(x)\} - H\{q_n^*(x)\}| + |H\{q_n^*(x)\} - H\{q(x)\}| \\ &\le \sup_{x\in\mathbb{R}} |H_n(x) - H(x)| + |H\{q_n^*(x)\} - H\{q(x)\}| \to 0 \end{aligned}$$

almost surely, and equivalently, for each $\alpha \in (0,1)$,

$$P\left([2\log(1/h)]^{1/2}\left[\|Y_{n,6}\|/C_{K,1}^{1/2} - d_n\right] \le q_n^*(1-\alpha)\right) \to 1-\alpha \quad \text{almost surely,}$$

which means the consistency of the bootstrap confidence bands (2.8). □

The estimator $\hat{f}_n$ determines the center of both bootstrap and asymptotic confidence bands (cf. equation (2.8) and Corollary 1 in [30]). However, the two types of bands differ in width. For the bootstrap confidence bands, factor

$$C_{K,1}^{1/2}\left[\frac{x}{[2\log(1/h)]^{1/2}} + d_n\right]$$

in the equation describing the half of the width of the asymptotic confidence bands is replaced by the quantile $q_{1-\alpha}^*$. The simulation studies described in the next section show the advantage of the bootstrap method over the asymptotic method for small data samples.

3. SIMULATION RESULTS

In this section we present results of applying the method discussed above in a simulation study. For comparison purposes we also include asymptotic confidence bands from [30]. The simulations were conducted in the R environment. All results are based on 1000 simulation runs. For the unknown probability density f, nine functions supported on $[0,1]$ were considered (all taken from [30]):

Decreasing	$B(1,3)$	$f(x) = 3(1-x)^2$,
Unimodal	$B(2,4)$	$f(x) \sim x(1-x)^3$,
Unimodal	$B(5,3)$	$f(x) \sim x^4(1-x)^2$,
Bimodal	BM1	$0.55 \cdot B(3,7) + 0.45 \cdot B(7,3)$,
Bimodal	BM2	$0.45 \cdot B(6,13) + 0.55 \cdot B(15,8)$,
Constant	Unif	$f(x) = 1$,
Increasing	$B(2,1)$	$f(x) = 2x$,
Triangular	TR	$f(x) = 4xI_{[0,0.5]} + 4(1-x)I_{(0.5,1]}$,
Step function	SF	$f(x) = 0.6I_{[0,1/3]} + 0.9I_{(1/3,0.75]} + 1.7I_{(0.75,1]}$,

where $B(\alpha,\beta)$ stands for the beta distribution. The first five functions satisfy the assumptions formulated in Section 2.2. The last four densities do not satisfy the assumptions and were included to check the performance of the confidence bands, when some of the conditions are violated. An additional difficulty with SF is that it is not continuous.

Given a density function f of the squared spheres radii, artificial data samples from the density g of the squared circles radii were generated with the algorithm described in [30]. Smoothed bootstrap samples were generated using the kernel function

$$\bar{K}_0(x) = (15/16)(1-x^2)^2 I_{[-1,1]}(x).$$

The bootstrap quantile of $\|Y_n^*\|$ was simulated each time on the basis of 1000 bootstrap samples. Asymptotic confidence bands were constructed in the same way as in the simulation studies in [30]. In particular, the kernel function

$$K_0(x) = (15/16)(1-x^2)^2 I_{[-1,1]}(x)$$

was used, while the smoothing parameter h was selected according to the procedure proposed in [30]. The same method was used to select h when constructing bootstrap bands, according to Corollary 2.2. The smoothing parameter η was arbitrarily determined as a fraction of the estimated parameter h ($\eta = 0.7h$ or $\eta = 0.9h$, see details later in this section). All confidence bands were constructed on the interval $[a,b] = [0.1, 0.9]$.

Tables 1 and 2 show the simulated coverage probabilities and mean confidence band areas for asymptotic and bootstrap 90%-bands, for all considered probability densities f, and for sample sizes $n = 500, 1000, 3000, 5000$. For each individual sample, two bands were constructed: asymptotic and bootstrap. Each time, the smoothing parameter h was selected based on the data, using the procedure proposed in [30], while the parameter η was assumed to be equal to $0.7h$.

Table 1

Simulated coverage probabilities and mean band areas for asymptotic and bootstrap 90%-confidence bands, for densities that satisfy the assumptions and for various sample sizes n. The smoothing parameter h was selected according to the method proposed in [30], and $\eta = 0.7h$. The approximate standard error for the simulated coverage probability equals 0.9%.

Density	n	Method			
		Asymptotic		Bootstrap	
		Coverage	Area	Coverage	Area
$B(1,3)$	500	74.5	0.99	80.0	1.07
	1000	81.0	0.74	86.5	0.79
	3000	87.4	0.44	90.8	0.48
	5000	88.6	0.34	91.2	0.38
$B(2,4)$	500	69.5	1.14	83.3	1.39
	1000	77.4	0.87	87.3	1.03
	3000	84.7	0.54	90.0	0.62
	5000	86.2	0.42	91.5	0.49
$B(5,3)$	500	68.8	1.26	75.1	1.35
	1000	76.3	1.00	79.5	1.10
	3000	83.6	0.71	87.1	0.75
	5000	86.3	0.58	89.8	0.64
BM1	500	82.3	1.31	82.8	1.33
	1000	85.5	1.05	86.4	1.13
	3000	86.8	0.68	87.5	0.67
	5000	89.1	0.56	89.7	0.59
BM2	500	28.1	1.00	59.0	1.39
	1000	65.3	1.02	81.1	1.18
	3000	52.5	0.71	79.0	0.89
	5000	48.5	0.62	75.5	0.78

For both asymptotic and bootstrap confidence bands, the results were almost always better for a larger sample size. The exceptions were the cases of BM2 and SF, which were discussed in this context in [30]. Moreover, when comparing the two types of bands, one can see that, regardless of the sample size, bands constructed using bootstrap methods for densities that satisfy the assumptions (results from Tab. 1) have the actual probability of coverage closer to the nominal, but are almost always wider.

Table 2

Similar to Table 1, but for densities that do not satisfy the assumptions.

Density	n	Method			
		Asymptotic		Bootstrap	
		Coverage	Area	Coverage	Area
Unif	500	82.6	1.33	82.3	1.31
	1000	89.3	1.08	89.3	1.07
	3000	92.4	0.69	92.3	0.68
	5000	91.2	0.56	92.3	0.56
$B(2,1)$	500	73.0	1.25	75.7	1.31
	1000	84.9	1.06	82.6	1.04
	3000	86.9	0.75	87.7	0.74
	5000	91.2	0.61	91.4	0.61
TR	500	77.1	1.28	79.1	1.30
	1000	78.8	1.00	81.0	0.99
	3000	88.9	0.68	89.5	0.67
	5000	88.0	0.55	90.1	0.56
SF	500	46.2	0.96	46.5	0.97
	1000	78.7	1.16	78.8	1.16
	3000	67.2	0.91	69.6	0.93
	5000	60.0	0.83	58.7	0.83

The results for densities that do not satisfy the assumptions (Tab. 2) are comparable for both types of bands. For small sample sizes, one can alternatively use larger values for the smoothing parameter η, which will make the bands wider and the coverage probability higher. This effect is illustrated by the results from Tables 3 and 4, where $\eta = 0.9h$ was taken for $n = 500, 1000$.

Table 3

Simulated coverage probabilities and mean band areas for asymptotic and bootstrap 90%-confidence bands, for densities that satisfy the assumptions and for smaller sample sizes $n = 500$ and $n = 1000$. The smoothing parameter h was selected according to the method proposed in [30], and $\eta = 0.9h$. The approximate standard error for the simulated coverage probability equals 0.9%.

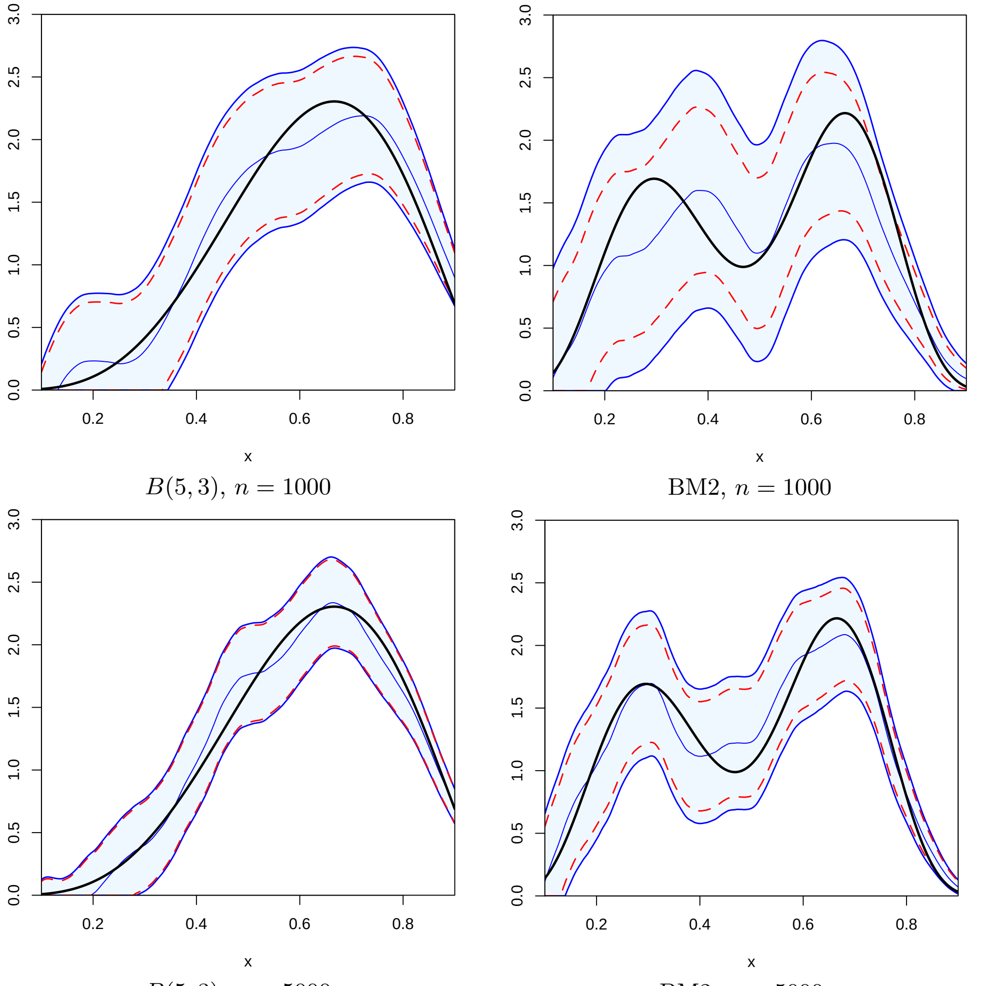
Density	n	Method			
		Asymptotic		Bootstrap	
		Coverage	Area	Coverage	Area
$B(1,3)$	500	73.5	0.97	86.3	1.15
	1000	82.5	0.74	91.1	0.87
$B(2,4)$	500	71.8	1.12	91.2	1.48
	1000	79.0	0.89	92.0	1.18
$B(5,3)$	500	69.4	1.26	78.8	1.32
	1000	76.0	1.00	86.0	1.05
BM1	500	81.4	1.32	87.2	1.40
	1000	84.8	1.06	88.2	1.11
BM2	500	29.0	1.01	80.0	1.67
	1000	62.8	1.02	89.6	1.38

Table 4

Similar to Table 3, but for densities that do not satisfy the assumptions.

Density	n	Method			
		Asymptotic		Bootstrap	
		Coverage	Area	Coverage	Area
Unif	500	82.4	1.33	86.2	1.36
	1000	88.2	1.08	91.2	1.11
$B(2,1)$	500	74.0	1.29	79.0	1.32
	1000	83.4	1.01	85.2	1.03
TR	500	76.8	1.28	85.4	1.36
	1000	80.0	1.01	86.7	1.05
SF	500	45.8	0.96	64.1	1.11
	1000	77.6	1.16	81.6	1.18

Figure 1 presents some typical examples of bootstrap and asymptotic confidence bands (in each case constructed on the basis of the same sample), which reflect the results presented in Tables 1–4. The parameter $\eta = 0.9h$ was selected for $n = 1000$ (as in the case of Tables 3 and 4), while $n = 5000$ corresponds to $\eta = 0.7h$ (as for the results from Tables 1 and 2).



$B(5,3)$, $n = 1000$ BM2, $n = 1000$

$B(5,3)$, $n = 5000$ BM2, $n = 5000$

Fig. 1. Estimate $\hat{f}_n$ along with associated 90% nominal coverage probability bootstrap confidence bands (——) and 90% nominal coverage probability asymptotic confidence bands (– – –), both types of bands constructed from the same data, for the densities $B(5,3)$ and BM2, and for two sample sizes $n = 1000$, $n = 5000$. The thick solid line represents the true function.

Acknowledgements

The author would like to thank Professor Zbigniew Szkutnik for his valuable comments and suggestions to improve the paper. The computational part of this research was supported in part by PL-Grid Infrastructure.

REFERENCES

[1] A. Antoniadis, J. Fan, I. Gijbels, *A wavelet method for unfolding sphere size distributions*, The Canadian Journal of Statistics **29** (2001), 251–268.

[2] P.J. Bickel, M. Rosenblatt, *On some global measures of the deviations of density function estimates*, The Annals of Statistics **1** (1973), 1071–1095.

[3] M. Birke, N. Bissantz, H. Holzmann, *Confidence bands for inverse regression models*, Inverse Problems **26** (2010), Article 115020.

[4] N. Bissantz, L. Dümbgen, H. Holzmann, A. Munk, *Non-parametric confidence bands in deconvolution density estimation*, Journal of the Royal Statistical Society: Series B **69** (2007), 483–506.

[5] N. Bissantz, H. Holzmann, *Statistical inference for inverse problems*, Inverse Problems **24** (2008), Article 034009.

[6] N. Bissantz, H. Holzmann, K. Proksch, *Confidence regions for images observed under the Radon transform*, Journal of Multivariate Analysis **128** (2014), 86–107.

[7] S.N. Chiu, D. Stoyan, W.S. Kendall, J. Mecke, *Stochastic Geometry and Its Applications*, Wiley, Chichester, 2013.

[8] L.M. Cruz-Orive, *Distribution-free estimation of sphere size distributions from slabs showing overprojections and truncations, with a review of previous methods*, J. Microscopy **131** (1983), 265–290.

[9] M. Csörgő, P. Révész, *Strong Approximations in Probability and Statistics*, Akadémiai Kiadó, Budapeszt, 1981.

[10] B. Ćmiel, Z. Szkutnik, J. Wojdyła, *Asymptotic confidence bands in the Spektor––Lord–Willis problem via kernel estimation of intensity derivative*, Electronic Journal of Statistics **12** (2018), 194–223.

[11] D. De Angelis, G.A. Young, *Smoothing the bootstrap*, International Statistical Review **60** (1992), 45–56.

[12] A. Delaigle, P. Hall, F. Jamshidi, *Confidence bands in nonparametric errors-in-variables regression*, Journal of the Royal Statistical Society: Series B **77** (2015), 149–169.

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

[14] B. Efron, *The Jackknife, the Bootstrap and Other Resarnpling Plans*, SIAM, Philadelphia, 1982.

[15] B. Efron, G. Gong, *A leisurely look at the bootstrap, the jackknife and crossvalidation*, The American Statistician **37** (1983), 36–48.

[16] E. Giné, R. Nickl, *Mathematical Foundations of Infinite-Dimensional Statistical Models*, Cambridge University Press, New York, 2016.

[17] G.K. Golubev, B.Y. Levit, *Asymptotically efficient estimation in the Wicksell problem*, Annals of Statistics **26** (1998), 2407–2419.

[18] P. Groeneboom, G. Jongbloed, *Nonparametric Estimation under Shape Constraints*, Cambridge University Press, New York, 2014.

[19] P. Hall, T.J. DiCiccio, J.P. Romano, *On smoothing and the bootstrap*, Annals of Statistics **17** (1989), 692–704.

[20] P. Hall, R.L. Smith, *The kernel method for unfolding sphere size distributions*, Journal of Computational Physics **74** (1988), 409–421.

[21] K. Kato, Y. Sasaki, *Uniform confidence bands in deconvolution with unknown error distribution*, Journal of Econometrics **207** (2018), 129–161.

[22] K. Kato, Y. Sasaki, *Uniform confidence bands for nonparametric errors-in-variables regression*, Journal of Econometrics **213** (2019), 516–555.

[23] K. Lounici, R. Nickl, *Global uniform risk bounds for wavelet deconvolution estimators*, Annals of Statistics **39** (2011), 201–231.

[24] K. Proksch, N. Bissantz, H. Dette, *Confidence bands for multivariate and time dependent inverse regression models*, Bernoulli **21** (2015), 144–175.

[25] B.W. Silverman, G.A. Young, *The bootstrap: to smooth or not to smooth?*, Biometrika **74** (1987), 469–479.

[26] W. Stute, *A law of the logarithm for kernel density estimators*, The Annals of Probability **10** (1982), 414–422.

[27] C.C. Taylor, *A new method for unfolding sphere size distributions*, Journal of Microscopy **132** (1983), 57–66.

[28] S.D. Wicksell, *The corpuscle problem: A mathematical study of a biometric problem*, Biometrika **17** (1925), 84–99.

[29] B.B. Winter, *Convergence rate of perturbed empirical distribution functions*, Journal of Applied Probability **16** (1979), 163–173.

[30] J. Wojdyła, Z. Szkutnik, *Nonparametric confidence bands in Wicksell's problem*, Statistica Sinica **28** (2018), 93—113.

Jakub Wojdyła
jwojdyla@agh.edu.pl
https://orcid.org/0000-0003-1892-8248

AGH University of Science and Technology
Faculty of Applied Mathematics
al. Mickiewicza 30, 30-059 Kraków, Poland

Received: March 15, 2021.
Accepted: September 4, 2021.