Statistica Sinica **28** (2018), 93-113
doi:https://doi.org/10.5705/ss.202015.0471

NONPARAMETRIC CONFIDENCE BANDS IN WICKSELL'S PROBLEM

Jakub Wojdyła and Zbigniew Szkutnik

AGH University of Science and Technology

Abstract: Although asymptotic nonparametric confidence bands have been constructed in the last decade in some inverse problems, like density deconvolution, inverse regression with a convolution operator, and regression with errors in variables, there seems to be no such construction for practically important inverse problems of stereology. Working with a kernel-type nonparametric estimator of the density of squared radii in the stereological Wicksell's problem, we partially fill this gap by constructing some corresponding asymptotic uniform confidence bands and an automatic bandwidth selection method, tuned to perform well in finite samples in terms of both area and coverage probability of the confidence bands. The performance of the new procedures is investigated in simulations and demonstrated with some astronomical data related to the M62 globular cluster.

Key words and phrases: Abel integral equation, ill-posed inverse problem, kernel methods, nonparametric curve estimation, stereology.

1. Introduction

Consider a population of spheres of random radii, randomly distributed in some three-dimensional opaque medium. The goal is to estimate the density of those spheres radii, when data are available only from a random plane slice through the medium. This problem, first posed by Wicksell (1925), became a classical problem in stereology. A rigorous treatment using marked point processes formalism, along with many applications in diverse areas, e.g., in biology, astronomy, geology and metallurgy, can be found, e.g., in Chiu et al. (2013), and in the references given there.

Throughout this paper, as in, e.g., Hall and Smith (1988), Golubev and Levit (1998), Antoniadis, Fan and Gijbels (2001) and others, we consider squared radii of both the unobserved spheres of interest and the observed circular sections, which is more convenient mathematically and sometimes also more natural to interpret. As noted by Hall and Smith (1988, p. 411), "the practical motivation is that the squared radius is proportional to the observed cross-sectional area,

which may be easier to measure than the radius." It is clear, however, that the results obtained for squared radii can easily be transformed back to the original problem.

More specifically, let f and g denote, respectively, the density of the squared spheres radii and the density of the observable squared circles radii. Then (see, e.g., Hall and Smith (1988) or Groeneboom and Jongbloed (2014, Sec. 4.1))

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

where $m = 2\int_0^\infty x^2 f(x^2)\, dx$ is the mean sphere radius. Equation (1.1), with temporarily fixed m, is a special case of the Abel integral equation and is related to fractional integration of order $1/2$ (see, e.g., Andersen and de Hoog (1990)). The solution, when it exists, is given by

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

and corresponds to half-differentiation of g.

In standard L_2-settings, the problem of unfolding f from g is ill-posed and requires some regularization. This is because the integral operator in (1.1) is compact, with its singular values approaching zero at the rate $i^{-1/2}$, and, consequently, the inverse operator in (1.2) is unbounded. Another way of looking at the difficulty of unfolding the balls radii distribution is to formulate the problem as that of estimating the cumulative distribution function, say F, rather than the density f. Then, see, e.g., Exercise 4.2 in Groeneboom and Jongbloed (2014),

$$F(x) = 1 - \frac{\int_x^\infty (z-x)^{-1/2} g(z)\, dz}{\int_0^\infty z^{-1/2} g(z)\, dz}.$$

A natural estimator of F, obtained via replacement of $g(z)\, dz$ with $dG_n(z)$, where G_n is the empirical distribution of the observed squared radii, is, however, neither monotone nor even bounded. This can naturally be handled with isotonization, which leads to attractive estimators of F as well as to asymptotic and bootstrap pointwise confidence intervals (see Groeneboom and Jongbloed (1995) and Sen and Woodroofe (2012)). In this paper, however, we concentrate on the classical Wicksell's problem of estimating the density function.

There are many different approaches to solving the Wicksell's problem. For a comprehensive overview of existing methods, see, e.g., Ripley (1981, Sec. 9.4) and Chiu et al. (2013, Sec. 10.4.2). With our main goal of constructing confidence bands for f and given available analytical techniques, we concentrate on kernel-type estimators. Taylor (1983) was the first to introduce a kernel-based approach

to this problem. He suggested estimating the density of interest by unfolding a kernel estimator of g, derived from the observed profiles. Hall and Smith (1988) investigated theoretical properties of Taylor's estimator and showed that it is pointwise optimal in the minimax sense. An alternative method was proposed by van Es and Hoogendoorn (1990). They postponed the kernel smoothing until after the inverse transformation step, but concluded that there is no reason to generally prefer one of the two possible orders of estimation and smoothing steps over the other. Golubev and Levit (1998) derived a kernel-based approach to estimating the distribution function of the squared spheres radii.

Confidence bands provide the most informative way of quantifying the accuracy of estimators in various problems of function estimation. Much effort has been put to the construction of confidence bands in direct problems, starting with the pioneering work of Bickel and Rosenblatt (1973), who constructed confidence bands for the density function of independent and identically distributed observations. Since then, their method has been further developed and applied to various setups. For reviews of papers on this topic, see Bissantz et al. (2007), Birke, Bissantz and Holzmann (2010) and Proksch, Bissantz and Dette (2015). Adaptivity issues in constructing confidence bands for a density were addressed by Giné and Nickl (2010).

Work on construction of nonparametric confidence bands in the inverse problem setup has started only recently. The first step in this direction seems to be the work by Bissantz et al. (2007), who constructed asymptotic and bootstrap confidence bands in ordinary smooth deconvolution problems. Since then, several related works have been published. For example, Bissantz and Holzmann (2008) and Lounici and Nickl (2011) restudied the construction of confidence bands in deconvolution density estimation problems. Birke, Bissantz and Holzmann (2010) provided confidence bands in a one-dimensional indirect regression model with a convolution operator. Methods for constructing confidence bands in a nonparametric errors-in-variables regression were developed by Delaigle, Hall and Jamshidi (2015). Recently, Proksch, Bissantz and Dette (2015) constructed confidence bands for the regression function in an inverse regression model with a convolution-type operator and with, for the first time, a multivariate predictor.

Stereological problems constitute an important subclass of inverse problems. However, despite their practical importance, confidence bands for densities in stereological problems have apparently not been considered so far. The purpose of this paper is to partially fill this gap by constructing asymptotic confidence bands in one of the most popular problems of stereology—the Wicksell's corpuscle

problem. To achieve this, we use a kernel-type estimator and, following Bickel and Rosenblatt (1973) and Bissantz et al. (2007), construct asymptotic confidence bands that are based on strong approximations and on a limit theorem for the supremum of a stationary Gaussian process. It should be stressed here that, although (1.1) resembles a convolution of f with the density $\psi(z) = 1/(2\sqrt{-z})$, $z \in [-1, 0)$, the Wicksell's problem is not a deconvolution problem, because the convoluted density is additionally lower truncated at zero and renormalized.

To be fully practicable, our method requires an algorithm for choosing the bandwidth. We propose a modification of the data-driven bandwidth selection procedure introduced by Bissantz et al. (2007), focusing on ensuring a good finite-sample performance in terms of both the coverage probability and the size of the resulting confidence bands.

The outline of the paper is as follows. The kernel-type estimator and the basic assumptions are presented in Sections 2.1 and 2.2, respectively. Section 2.3 contains asymptotic results needed for construction of confidence bands. Section 3 is devoted to a numerical implementation of our procedures (with a data-driven choice of the bandwidth) and reports on some Monte Carlo experiments. In Section 4, we illustrate our approach on some data in a closely related globular cluster problem of astronomy. Proofs are given in the Appendix and in the online Supplement.

2. Confidence Bands

In this section, the main result of this article is presented: the construction of asymptotic confidence bands for f on a compact subset of its support.

2.1. Central estimator

The confidence bands are constructed around the Hall and Smith (1988) version of the kernel-type estimator originally proposed by Taylor (1983). In order to estimate the density of the squared spheres radii from the observed squared radii $Y_1^2, \ldots, Y_n^2$ of circular profiles, the ordinary kernel estimator $\hat{g}_n$ of the density g,

$$\hat{g}_n(y) = \frac{1}{nh} \sum_{i=1}^{n} K_0\left(\frac{y - Y_i^2}{h}\right), \quad y \geq 0,$$

is substituted to (1.2), which gives an estimator of f of the form

$$f_n(x) = \frac{-2m}{\pi} \frac{d}{dx} \int_x^\infty (y - x)^{-1/2} \hat{g}_n(y)\, dy, \quad x \geq 0.$$

For a regular kernel K_0, it can be transformed to an easily computable kernel form

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

with the kernel

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

The estimator defined in this way depends on the parameter m, which is usually unknown in practice. One possible remedy to that is to confine oneself to estimating f/m instead of f (cf., e.g., van Es and Hoogendoorn (1990)). If necessary, the estimator of f/m can be rescaled to a density. In our theoretical results and simulations, however, m is replaced with an appropriate estimator $\hat{m}$.

Hall and Smith (1988) showed that, with sufficiently regular kernel, Taylor's estimator is pointwise optimal in the minimax sense for the class of k-times boundedly differentiable densities f. As the upper bounds are uniform in x, they can also be used to produce upper bounds for the L_2-risk. Those coincide with known lower bounds of the order of $n^{-2k/(2k+2)}$, which means that the estimator is rate minimax also in the global L_2-sense. For the construction of confidence bands, however, this aspect is only of secondary importance, and is not discussed here in any more detail.

2.2. Assumptions

In the Wicksell's corpuscle problem, it is typically assumed that the density f (and hence also g) has a bounded support $[0, R]$. Without loss of generality, we assume here $R = 1$. For reasons explained, e.g., by van Es and Hoogendoorn (1990), the density f cannot be reliably estimated in a vicinity of zero. Also, because of technical reasons, we need a positive lower bound for g, which means that arguments in a neighbourhood of one must be excluded. Therefore, the confidence bands will be constructed on an interval $[a, b] \subset [0, 1]$, with $a > 0$ and $b < 1$.

In the sequel, we impose the following regularity conditions.

Assumptions on the kernel:

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

(1b) 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

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

Assumptions on the problem:

(2a) The density g is bounded away from zero on $[a, b]$ and $g^{1/2}$ is differentiable with bounded derivative on $[0, 1]$.

(2b) For k as in assumption (1a), f is $(k-1)$-times continuously differentiable in $[0, 1]$ and there exists bounded $f^{(k)}$ in $(0, 1)$.

The biweight kernel $K_0(x) = (15/16)(1-x^2)^2 I_{[-1,1]}(x)$, with I denoting the indicator function, is an example of a kernel function that satisfies assumptions (1a) and (1b) with $k = 2$. This function is used in our simulation studies in Section 3 and in the real data example in Section 4. Verification of assumption (2a) for some given f can be difficult, because it is expressed in terms of g, related to f at (1.1). By a direct calculation, this can be done for, e.g., densities of the beta distribution. For more details, see Section 3.

2.3. Main results

Adapting the Bickel-Rosenblatt methodology, we start with investigating the limiting distribution of the supremum of the process

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

where the estimator f_n is defined in (2.1). Denote with $\|\cdot\|$ the sup-norm on the chosen interval $[a, b]$. The following theorem is proved in the Appendix.

Theorem 1. *Under assumptions* (1b) *and* (2a)*, if* $h \to 0$ *and* $nh/(\log n)^3 \to \infty$*, then, for each* $x \in \mathbb{R}$*,*

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

where

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

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

If m and $g(t)$ with $t \in [a, b]$ were known, confidence bands for $E(f_n)$ would directly follow from Theorem 1. (We will propose appropriate estimators for m and $g(t)$ later.) To construct confidence bands for f, one has to control the bias of f_n. We achieve that by undersmoothing, as in Bissantz et al. (2007), Bissantz and Birke (2009), Birke, Bissantz and Holzmann (2010), and Proksch, Bissantz and Dette (2015) (cf. also Giné and Nickl (2016, Sec. 6.4.2)). This means, in our case, choosing a bandwidth that converges to zero faster than $n^{-1/(2k+2)}$, which is the rate of the optimal bandwidth in the mean squared error sense (Hall and Smith (1988)). The choice of sufficiently small bandwidths assures the same limiting behaviour of $f_n(t) - E\{f_n(t)\}$ and $f_n(t) - f(t)$ (see the proof of Corollary 1 in the online Supplement). There are also some theoretical arguments implying that undersmoothing should be a preferable method when constructing confidence intervals (see Hall (1992)). For an alternative approach to controlling the bias, see, e.g., Eubank and Speckman (1993), where the bias is explicitly estimated and substracted.

Hall and Smith (1988) proved that

$$E\{f_n(x)\} = \int K_0(z) f(x - hz)\, dz.$$

Therefore, the mean of f_n has the same form as the mean of the ordinary kernel estimator with the kernel K_0. Consequently, standard reasoning based on the Taylor expansion of f (see, e.g., Silverman (1986, Sec. 3.3.1)) shows that, under our assumptions,

$$|E\{f_n(x)\} - f(x)| = O(h^k), \tag{2.3}$$

uniformly in $x \in [a, b]$. To continue the construction, one needs to estimate the unknown density g of the observations. We assume that this is done using an estimator $\tilde{g}_n$, not necessarily equal to $\hat{g}_n$, that satisfies

$$\|\tilde{g}_n - g\| = o_p\left(\frac{1}{\log(1/h)}\right), \tag{2.4}$$

where h is the bandwidth chosen for the construction of the estimator f_n of f. Further, one needs to estimate the unknown mean m. Assume that this is done with an estimator $\hat{m}$ such that, for all $\varepsilon > 0$,

$$\hat{m} - m = O_p(n^{-1/2+\varepsilon}). \tag{2.5}$$

Examples of such estimators $\tilde{g}_n$ and $\hat{m}$ are given at the end of this section.

Denote with $\hat{f}_n$ the estimator obtained from f_n by replacing m in the definition (2.1) with $\hat{m}$,

$$\hat{f}_n(x) = \frac{-2\hat{m}}{nh^{3/2}\pi} \sum_{i=1}^{n} K\left(\frac{x - Y_i^2}{h}\right). \tag{2.6}$$

The confidence bands constructed in this way have the form given in the following corollary, proved in the online Supplement.

Corollary 1. *Let $\tilde{g}_n$ be an estimator of g satisfying* (2.4) *and let $\hat{m}$ be an estimator of m satisfying* (2.5)*. Under assumptions* (1a), (1b), (2a)*, and* (2b)*, if $h \to 0$ in such a way that $n^{-1+\delta}\log(1/h) = O(1)$, $n^{\delta}h\log(1/h)^{1/2} = O(1)$ for some $\delta > 0$, $n^{1/2}h^{k+1}\log(1/h)^{1/2} \to 0$, and $nh^2/\log(1/h) \to \infty$, then, for each $x \in \mathbb{R}$,*

$$P(\hat{f}_n(t) - b_n(t,x) \le f(t) \le \hat{f}_n(t) + b_n(t,x) \text{ for all } t \in [a,b]) \to \exp\{-2\exp(-x)\}, \tag{2.7}$$

where

$$b_n(t,x) = \frac{2\hat{m}\tilde{g}_n(t)^{1/2}C_{K,1}^{1/2}}{n^{1/2}h\pi}\left[\frac{x}{[2\log(1/h)]^{1/2}} + d_n\right],$$

and the constants $C_{K,1}$ and d_n are defined in Theorem 1*.*

The assumptions imposed above on the rate of convergence of h to zero can be met simultaneously. For example, for all $\delta \in (0, 1/2)$, a bandwidth of the form $h = n^{-\gamma}$, with any $\gamma \in (\max\{\delta, 1/(2k+2)\}, 1/2)$, satisfies them. The width of the confidence bands is of the order of $\left[\log(1/h)/(nh^2)\right]^{1/2}$, and hence the condition $nh^2/\log(1/h) \to \infty$ ensures that the width converges to zero as $n \to \infty$.

With $h = n^{-\gamma}$, the convergence-in-probability condition (2.4) reduces to $\|\tilde{g}_n - g\| = o_p(1/\log n)$, which is satisfied by kernel estimators in typical cases, in which $E\|\tilde{g}_n - g\|$ decays as $n^{-\beta}$, with some $\beta > 0$ (cf., e.g., Dony and Einmahl (2006)).

Condition (2.5) is satisfied, for example, by the estimator

$$\hat{m} = \frac{n\pi}{2}\left(\sum_{i=1}^{n} Y_i^{-1}\right)^{-1}, \tag{2.8}$$

if the distribution of the spheres radii has finite variance (Ripley (1981, Sec. 9.4); Hall and Smith (1988)).

3. Implementation and Simulations

Finite sample performance of the proposed asymptotic confidence bands and of data-driven bandwidth selection algorithms was investigated in Monte Carlo experiments. The simulations were conducted in the R environment, version

3.2.0. All results are based on $1,000$ simulation runs. For the unknown probability density f, nine functions supported on $[0,1]$ were considered:

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 (taken from Dudek and Szkutnik (2008)) 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 following algorithm. Pairs (R^2, Z) of independent random variables were generated: R^2 from the density f and a distance Z from the ball center to the slicing plane, uniformly distributed on $[0,1]$. For $Z > R$ the points were dropped; otherwise, $R^2 - Z^2$ were taken as the observed squared circles radii.

$B(\alpha,\beta)$-distributed random numbers were computed with the function *rbeta()* from the R package *stats*. The variates from densities TR and SF were computed from the uniform ones by inversion of the corresponding cumulative distribution function. All uniform variates were generated using the function *runif()* from the R package *stats*.

In order to construct the confidence bands (2.7), we first estimated f with the estimator (2.6), with K obtained from the biweight kernel $K_0(x) = (15/16)(1-x^2)^2 I_{[-1,1]}(x)$ according to (2.2), with a data-driven bandwidth h chosen as described in the next subsection, and with the estimator $\hat{m}$ defined at (2.8). The Epanechnikov kernel, for which K has a simpler form, does not satisfy assumption (1b), and for that reason it is not used here; nevertheless, it was tried in simulations and performed comparably to the biweight kernel. As $\tilde{g}_n$, we tried kernel estimators with several kernels and several bandwidth choice methods. Extensive simulations, not reported here, showed that all of them produced similar results.

Therefore, we finally used the computationally cheap estimator implemented in the R function *density()* with a Gaussian kernel and a simple rule-of-thumb for choosing the bandwidth, implemented in the function *bw.nrd0()* from the R package *stats*. Finally, confidence bands on the interval $[a, b] = [0.1, 0.9]$ were constructed according to Corollary 1.

3.1. Bandwidth selection

It is known that bandwidth selection is crucial for the quality of nonparametric confidence bands (Bissantz et al. (2007); Bissantz and Holzmann (2008); Birke, Bissantz and Holzmann (2010)).

Figure 1 presents simulated coverage probabilities and average areas of nominal 90% confidence bands as functions of the bandwidth h for the true functions $B(5, 3)$ and BM1, and for sample size $n = 5{,}000$. The area of the confidence bands has been normalized such that its maximum value always equals one. The results are similar to those obtained previously for density deconvolution (Bissantz et al. (2007, Fig. 1)) and for inverse regression model (Birke, Bissantz and Holzmann (2010, Fig. 1)). For proper determination of the confidence bands, the bandwidth should be chosen somewhat smaller than the L_∞-optimal bandwidth, which corresponds to the minimum of the dotted curve (representing the mean L_∞-error $\|f_{n,h} - f\|$ as a function of h). This finite sample undersmoothing in the L_∞-sense corresponds to the asymptotic undersmoothing in the L_2-sense in Corollary 1 (cf. also the discussion directly following Theorem 1).

In practice, however, the true function f, used to produce the dotted curves, is not known and the estimation of the L_∞-optimal bandwidth is not straightforward. Standard data-driven bandwidth selection procedures for kernel density estimators usually tend to produce oversmoothed estimators, and cannot be used directly when constructing uniform confidence intervals. Bissantz et al. (2007) proposed a L_∞-based bandwidth selector for the density deconvolution problem, which was later successfully applied also by Bissantz and Holzmann (2008), Birke, Bissantz and Holzmann (2010) and Proksch, Bissantz and Dette (2015).

The algorithm of Bissantz et al. (2007) is based on the practical observation that, in the deconvolution problem studied by them, "the bandwidth, where $d^{(\infty)}_{j-1,j}$ changes its slope suddenly, is a good indicator of the bandwidth which minimizes the L_∞-distance between the true function and the estimator", where $d^{(\infty)}_{j-1,j}$ denotes the L_∞-distance between estimators $\hat{f}_{n,h_{j-1}}$ and $\hat{f}_{n,h_j}$ for two adjacent bandwidths h_{j-1}, h_j on a sufficiently dense grid (cf. Figures 2, 3 in Bissantz et al. (2007)). They concluded that, in the density deconvolution problem, such

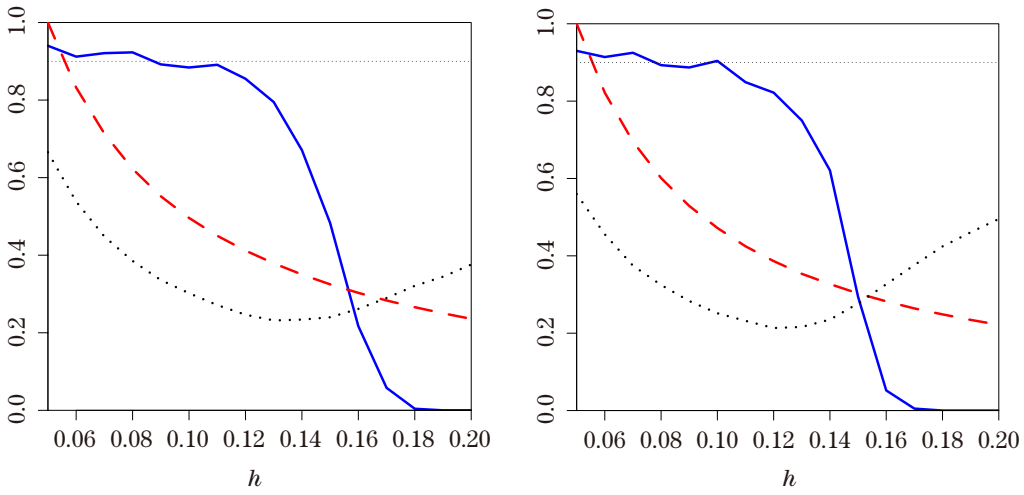

Figure 1. Simulated coverage probability (———) and normalized average area (– – –) of asymptotic confidence bands with a nominal coverage probability of 90% for the densities $B(5,3)$ (left) and BM1 (right) for 5,000 observations. The dotted curves (· · · · · ·) represent the mean L_∞ estimation errors and the horizontal dotted lines indicate the nominal coverage probability.

h_j works satisfactorily as a data-driven bandwidth in finite samples, in spite of theoretically motivated need for asymptotic undersmoothing.

Our preliminary simulations showed, however, that this algorithm, when applied to Wicksell's problem, often selects too small bandwidths and, as a result, the corresponding confidence bands are too wide. This usually happens for unimodal or flat densities. We thus had to modify the original algorithm.

Figure 2 shows, for four different samples, the plots of the L_∞ estimation errors for the densities $B(5,3)$ and BM1. For the same data sets, Figure 3 shows the L_∞-distances $d^{(\infty)}_{j,j+1}$ as function of the bandwidths computed on the grid $h_j = h_0 j/20$, for $j = 1, \ldots, 20$ and $h_0 = 0.2$. Figures 2 and 3 suggest that in our problem the smallest h_j, for which $d^{(\infty)}_{j,j+1} < d^{(\infty)}_{j+1,j+2}$, say h^*, is usually somewhat smaller than the bandwidth that minimizes the estimation error $\|\hat{f}_{n,h} - f\|$ and may possibly be a good candidate for a data-driven bandwidth in our problem. This conjecture was further confirmed in our extensive simulation studies (not reported here), in which our proposal was compared with the original algorithm of Bissantz et al. (2007). Except for BM2 and SF, the coverage probabilities for the confidence bands were satisfactory for both algorithms. However, the areas of confidence bands produced with our h^* tended to be significantly smaller. For example, for nominal 90% coverage probability and $n = 5,000$, the average

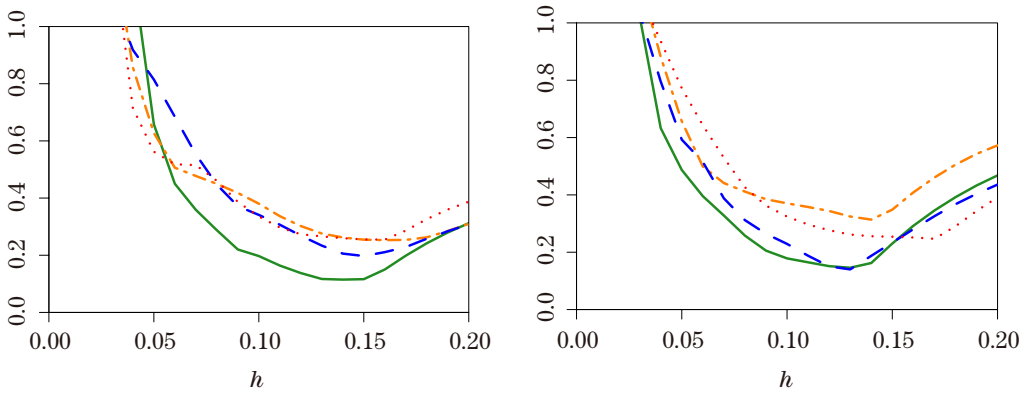

Figure 2. The L_∞ estimation error for densities $B(5,3)$ (left) and BM1 (right) as function of the bandwidth h, for four different samples of size 5,000.

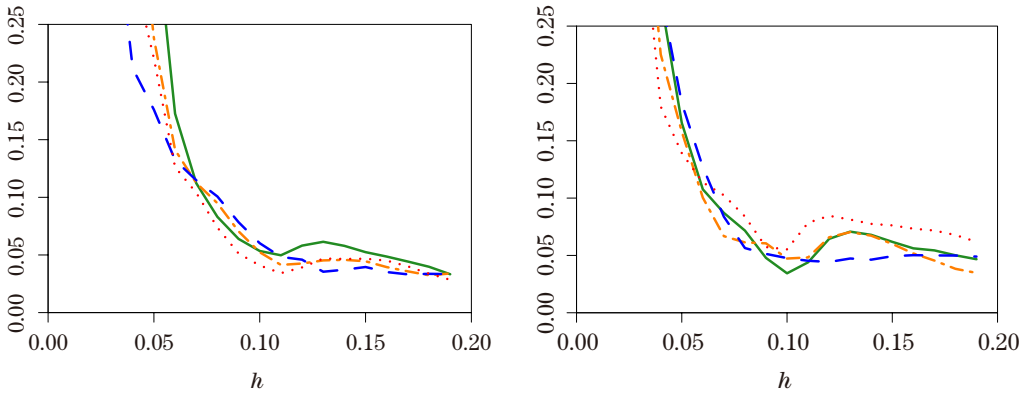

Figure 3. The L_∞-distance $d_{j,j+1}^{(\infty)}$ between the estimates $\hat{f}_{n,h_j}$ and $\hat{f}_{n,h_{j+1}}$ of $B(5,3)$ (left) and BM1 (right) as function of the bandwidths h_j, using the same curve style and the same samples as in Figure 2.

reduction of the areas with respect to those obtained with bandwidth selected with the method of Bissantz et al. (2007) amounted to 50.6%, 38.1%, 7.8%, 27.4%, 32.0%, 8.9%, and 17.3%, respectively, for $B(1,3)\div$BM1 and Unif$\div$TR.

Although the existence of h^* is not generally guaranteed, it existed in almost all cases studied in our simulations (96.7% for $B(5,3)$, 94.8% for BM2, and 100% in all remaining cases, with $n = 5,000$, $h_0 = 0.2$, $J = 20$, and $1,000$ runs for each f). Moreover, it was always tightly related to the bandwidth, say h_∞, optimal in the L_∞ error sense. For example, for $B(5,3)$ and $n = 5,000$, the mean value of the quotient h^*/h_∞ amounted to 0.76, with the variance of its distribution equal to 0.03.

Taking those observations into account, we suggest the following procedure for a data-driven bandwidth selection.

1. Choose some oversmoothing pilot bandwidth h_0.

2. If the density of interest is expected to be unimodal or arbitrary, but rather flat, compute $\hat{f}_{n,h_j}$ for a grid of J-values $h_j = h_0 j/J$, $j = 1, \ldots, J$, with $J \approx 20$ and choose the bandwidth as the smallest h_j such that $d^{(\infty)}_{j,j+1} < d^{(\infty)}_{j+1,j+2}$, if such h_j does exist. Otherwise, select the bandwidth using the algorithm from Bissantz et al. (2007), with the pilot bandwidth h_0, $J \approx 20$, and $\tau \approx 2$.

When selecting h_0, it is important to guarantee it is oversmoothing. In all simulations reported in this article, $h_0 = 0.2$ was used and proved satisfactory for all sample sizes and all densities. The same holds true for the selected values of J and τ. The impact of varying parameters on the selected bandwidth is discussed in detail by Bissantz et al. (2007).

In simulation studies reported in this section, the BM2 and SF functions were treated as "not expected to be unimodal or flat", which means that the original method of Bissantz et al. (2007) was used in those cases (with $J = 20$ and $\tau = 2.1$).

3.2. Simulation results

Tables 1 and 2 show the simulated coverage probabilities and the confidence band areas for all considered probability densities f, for sample sizes $n = 3{,}000$, 5,000, 7,000 and for three levels of nominal coverage probability. Except for BM2 and SF, the confidence bands perform reasonably well with respect to the coverage probabilities. For bimodal and relatively rapidly changing BM2, the procedure properly reconstructs the shape but has problems with covering the true function, especially at the right tail, which is close to zero. Here BM1, also bimodal but with less pronounced modes, poses no problems. The problem with SF is a rather obvious consequence of the lack of continuity. The results in Table 2 indicate that the proposed procedure may perform well even when not all of the assumptions are satisfied, provided the estimated function is continuous. Figure 4 shows some typical examples of 80% and 95% confidence bands for $n = 5{,}000$, obtained with h chosen with the procedure described in the previous section.

The estimator (2.8) of m was used in the simulation studies described in this section. In additional simulations, not reported in detail here, three other estimates of m were investigated. One of them was $\int_0^1 (1 - \hat{F}_n(x))\, dx$, where $\hat{F}_n$ is

Table 1. Simulated coverage probabilities and mean confidence band areas for densities that satisfy the assumptions and for various sample sizes n. Approximate standard errors for simulated coverage probabilities are 1.3%, 0.9%, and 0.7%, respectively, for nominal coverage 80%, 90%, and 95%.

		Nominal coverage					
		80%		90%		95%	
Density	n	Coverage	Area	Coverage	Area	Coverage	Area
$B(1,3)$	3,000	74.6	0.39	86.6	0.43	94.0	0.48
	5,000	74.7	0.31	89.4	0.35	94.5	0.38
	7,000	76.5	0.26	89.8	0.30	95.4	0.33
$B(2,4)$	3,000	71.0	0.47	85.1	0.54	90.3	0.60
	5,000	72.1	0.38	85.0	0.43	92.2	0.49
	7,000	70.6	0.33	85.1	0.37	91.0	0.42
$B(5,3)$	3,000	73.7	0.62	86.9	0.70	92.8	0.77
	5,000	73.7	0.51	86.2	0.57	92.5	0.64
	7,000	73.5	0.44	88.0	0.50	94.7	0.55
BM1	3,000	76.1	0.61	86.2	0.69	93.2	0.76
	5,000	77.2	0.48	88.1	0.55	93.4	0.61
	7,000	72.3	0.42	86.4	0.48	93.5	0.51
BM2	3,000	36.8	0.63	50.9	0.71	63.2	0.79
	5,000	34.9	0.55	47.6	0.61	60.1	0.68
	7,000	36.6	0.50	51.1	0.56	59.7	0.62

Table 2. Similar to Table 1, but for densities that do not satisfy the assumptions.

		Nominal coverage					
		80%		90%		95%	
Density	n	Coverage	Area	Coverage	Area	Coverage	Area
Unif	3,000	82.3	0.61	91.7	0.69	95.7	0.76
	5,000	82.8	0.49	92.8	0.55	97.4	0.61
	7,000	81.3	0.41	92.0	0.47	96.6	0.52
$B(2,1)$	3,000	76.9	0.65	87.3	0.75	94.9	0.82
	5,000	82.8	0.54	90.7	0.61	95.3	0.67
	7,000	81.4	0.47	92.7	0.53	95.8	0.59
TR	3,000	75.8	0.60	87.9	0.69	93.8	0.76
	5,000	77.1	0.49	88.5	0.55	95.3	0.61
	7,000	75.4	0.42	88.2	0.48	94.3	0.53
SF	3,000	47.1	0.82	66.9	0.92	79.1	1.01
	5,000	40.8	0.74	60.2	0.82	74.6	0.90
	7,000	34.6	0.68	54.4	0.77	67.1	0.83

the isotonic estimator of F proposed by Groeneboom and Jongbloed (1995) (see also Sen and Woodroofe (2012)). This gave results similar to those obtained with the estimator (2.8). Norming of the estimator of f/m and norming of its positive

part (both on $[0,1]$) worked, however, considerably worse than plugging-in the standard estimator (2.8).

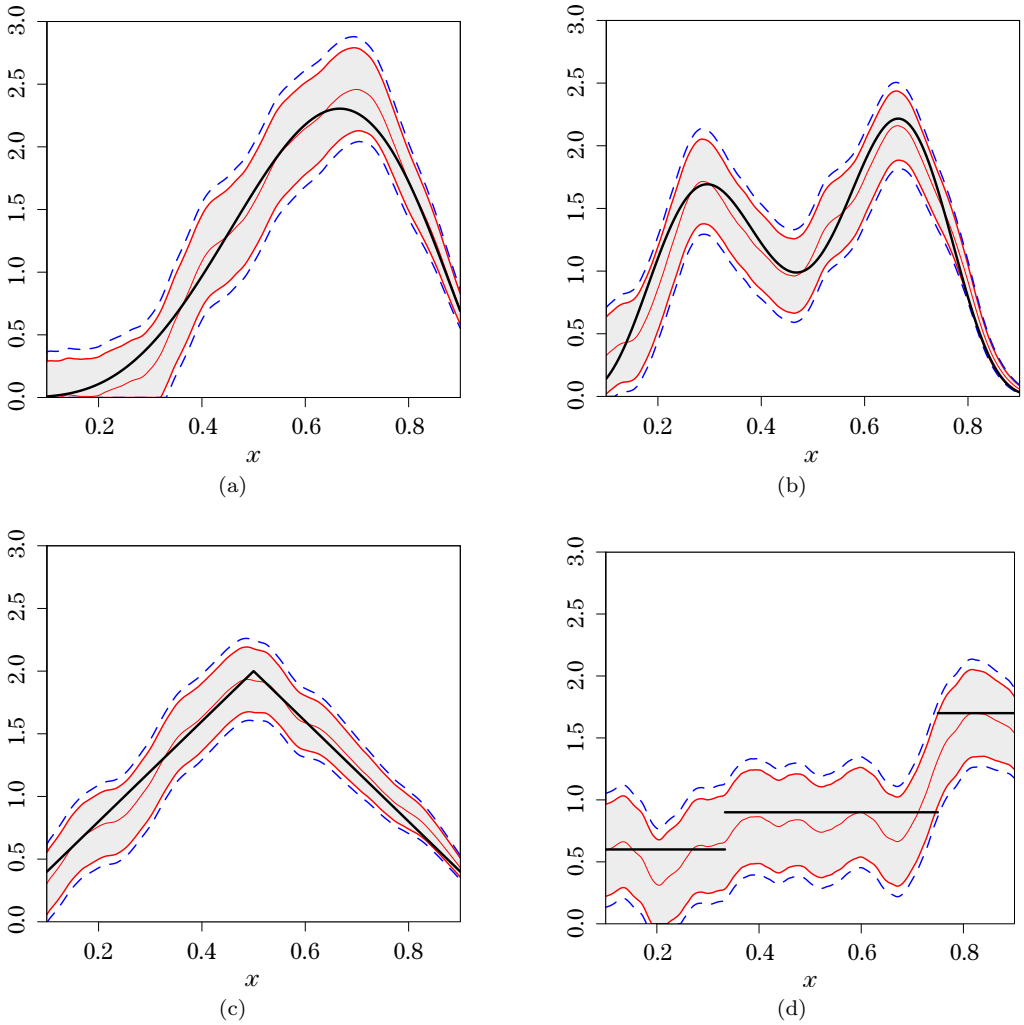

Figure 4. Estimate $\hat{f}_n$ along with associated 80% (——) and 95% (– – –) nominal coverage probability confidence bands of the densities (a) $B(5,3)$, (b) BM2, (c) TR, (d) SF. In all cases the sample size is $n = 5{,}000$ and the thick solid line represents the true function.

4. A Real Data Example

Let (X_1, X_2, X_3) denote the random position of a star in a spherically symmetric subsystem of a galaxy, with a density of the form $\rho(x_1^2+x_2^2+x_3^2)$. When the

star is observed through a telescope, only the projected stellar position (X_1, X_2) can be observed. Let f_1 and g_1 denote, respectively, the density of the squared distance $X_1^2 + X_2^2 + X_3^2$ of the star to the center of the cluster and the density of the squared distance $X_1^2 + X_2^2$ of the projected position of the star to the center of the projection plane. Then, it follows from Sen and Woodroofe (2012) that the relationship between f_1 and g_1 has the form

$$f_1(x) = \frac{-2\sqrt{x}}{\pi} \frac{d}{dx} \int_x^\infty (y-x)^{-1/2} g_1(y)\, dy, \quad x \geq 0,$$

which is almost identical to (1.2), considered in previous sections, but with $\sqrt{x}$ in place of the unknown m. This makes the astronomical problem slightly easier than Wicksell's problem. Obvious modifications of the result formulated in Corollary 1 give the confidence band for f_1:

$$\bar{f}_n(t) - \bar{b}_n(t,x) \leq f_1(t) \leq \bar{f}_n(t) + \bar{b}_n(t,x), \quad t \in [a,b],$$

where

$$\bar{f}_n(t) = \frac{-2\sqrt{t}}{nh^{3/2}\pi} \sum_{i=1}^n K\left(\frac{t - Z_i^2}{h}\right),$$

$$\bar{b}_n(t,x) = \frac{2\sqrt{t}\bar{g}_n(t)^{1/2} C_{K,1}^{1/2}}{n^{1/2}h\pi} \left[\frac{x}{[2\log(1/h)]^{1/2}} + d_n\right],$$

$Z_1^2, \ldots, Z_n^2$ are independent observations of $X_1^2 + X_2^2$, and $\bar{g}_n$ is a suitable estimator of g_1.

Globular clusters—compact, tightly bound by gravity groups of hundreds of thousands of stars, with the highest concentration of stars toward their centres—provide a good example of (approximately) symmetrical systems of stars. The study of globular clusters orbiting the core of the Milky Way can provide us with important clues about the evolution of the Galaxy (see, e.g., Alonso-García et al. (2012); Sen and Woodroofe (2012)).

Our analysis is based on 5,000 measurements of the projected positions of stars in the inner core of the globular cluster M62 (courtesy of B. Sen), the subset of which (of size 2,000) was previously used by Sen and Woodroofe (2012) for pointwise estimation of the distribution function of the squared distances in this cluster. Figure 5 shows the estimators and 95% confidence bands for the density of the squared distances of the stars of the cluster to its centre, obtained from the whole sample and from the subsample of 2,000 observations. The interval was rescaled to $[0,1]$ and the bands were constructed on $[0.1, 0.9]$ using the procedures proposed here. The result looks consistent with the estimated distribution

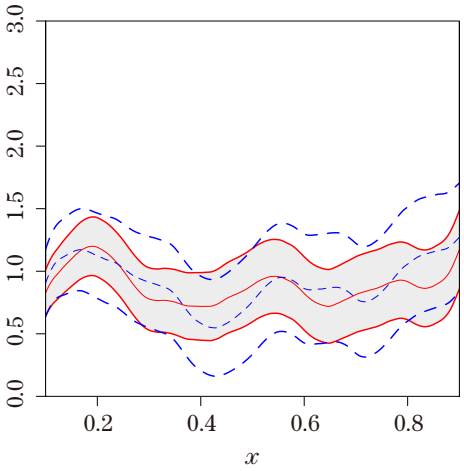

Figure 5. Estimates and associated 95% nominal coverage probability confidence bands for the density of the squared distances to the centre of the globular cluster M62, obtained from 2,000 (– – –) and 5,000 (———) observations.

function from Sen and Woodroofe (2012) and, being statistically consistent in the interval $[0.1, 0.9]$ with an approximately constant function, suggests a density of stars sharply growing towards the centre of the cluster.

Supplementary Materials

The online supplement contains proofs of Corollary 1 and of Lemma 1, formulated in the Appendix.

Acknowledgment

This work was partially supported by the Polish Ministry of Science and Education via an AGH local grant 10.420.03. The authors gratefully acknowledge the support from Professor Bothisattva Sen, who kindly provided them with astronomical data on the M62 globular cluster, previously used in his studies of a similar problem. The computational part of this research was supported in part by PL-Grid Infrastructure. The authors thank the anonymous reviewers for helpful suggestions of improvements.

Appendix: Proof of Theorem 1

The general idea of the proof is similar to that of the proofs of Theorem 3.1 of Bickel and Rosenblatt (1973) and of Theorem 1 of Bissantz et al. (2007). The

Hungarian embedding we employ in this context is well described in Giné and Nickl (2016, Sec. 5.1.3). The basic idea of the proof is to approximate the process Y_n with a Gaussian process that does not depend in any way on the true density function f. It is derived in several steps.

Let G denote the distribution function of the observed squared radii of the circular profiles, and let $\alpha_n^G(t) = n^{1/2}[G_n(t) - G(t)]$ be the corresponding empirical process, where G_n is the empirical distribution function based on $Y_1^2, \ldots, Y_n^2$. The Komlós-Major-Tusnády approximation (see, e.g., Theorem 4.4.1 in Csörgő and Révész (1981)) gives the existence of a sequence of Brownian bridges B_n such that

$$\sup_{t\in\mathbb{R}} |\alpha_n^G(t) - B_n\{G(t)\}| = O_p(n^{-1/2}\log n), \quad n \to \infty, \tag{A.1}$$

and $B_n(t) = W_n(t) - tW_n(1)$, $t \in [0,1]$, where W_n are standard Wiener processes.

Writing Y_n as a Stieltjes integral

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

and integrating by parts (due to assumption (1b), K' is absolutely integrable in $[0,1]$), one obtains

$$Y_n(t) = \frac{h^{-3/2}}{g(t)^{1/2}} \int_0^1 K'\left(\frac{t-x}{h}\right) \alpha_n^G(x)\, dx. \tag{A.2}$$

Consider two processes that approximate Y_n:

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

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

with B_n and W_n as in (A.1). Integrating by parts the stochastic integrals (cf. Corollaries 8.5 and 8.7 in Klebaner (2005)) and using the fact that K has bounded variation, since $\int |K'(x)|\, dx < \infty$ from assumption (1b), one obtains

$$Y_{n,0}(t) = \frac{h^{-3/2}}{g(t)^{1/2}} \int_0^1 K'\left(\frac{t-x}{h}\right) B_n\{G(x)\}\, dx. \tag{A.3}$$

From (A.2) and (A.3), substituting $u = (t-x)/h$, one obtains

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

It follows from assumption (2a), (A.1), and assumption (1b) that, as $n \to \infty$,

$$\|Y_n - Y_{n,0}\| = O_p(n^{-1/2}h^{-1/2}\log n). \tag{A.4}$$

Further, one has

$$|Y_{n,0}(t) - Y_{n,1}(t)| \le \frac{h^{-1/2}}{g(t)^{1/2}} \sup_{x\in[0,1]} |g(x)||W_n(1)| \left| \int_0^1 K\left(\frac{t-x}{h}\right)\, dx \right|.$$

Substituting $u = (t-x)/h$ and using assumptions (2a), and (1b), one obtains, as $n \to \infty$,

$$\|Y_{n,0} - Y_{n,1}\| = O_p(h^{1/2}). \tag{A.5}$$

For further approximation steps, define the processes:

$$Y_{n,2}(t) = \frac{h^{-1/2}}{g(t)^{1/2}} \int_0^1 K\left(\frac{t-x}{h}\right) g(x)^{1/2}\, dW(x), \quad t \in [a,b],$$
$$Y_{n,3}(t) = h^{-1/2} \int K\left(\frac{t-x}{h}\right)\, dW(x), \quad t \in [a,b],$$

where W is a two-sided Wiener process on $\mathbb{R}$.

The processes $Y_{n,0}$, $Y_{n,1}$, $Y_{n,2}$, $Y_{n,3}$ are well defined (see, e.g., Klebaner (2005, Chap. 8)), since the corresponding integrands are square integrable on $\mathbb{R}$. For convenience, suppose that the sample paths of all the processes defined above belong to $D[a,b]$, the space of cadlag functions on $[a,b]$.

Integration by parts and the substitution $u = (t-x)/h$ give

$$Y_{n,2}(t) = \frac{h^{-1/2}}{g(t)^{1/2}} \int_{(t-1)/h}^{t/h} \left[K'(u) g(t-hu)^{1/2} - hK(u) \frac{g'(t-hu)}{2g(t-hu)^{1/2}} \right] W(t-hu)\, du. \tag{A.6}$$

The following lemma is proved in the online Supplement.

Lemma 1. *Under assumptions* (1b) *and* (2a)*, with* $\alpha > 0$ *as in assumption* (1b)*, as* $n \to \infty$*,*

$$\|Y_{n,2} - Y_{n,3}\| = O_p(h^{\min\{\alpha/2, 1/2\}}).$$

The process $\{Y_{n,3}(t) : a \le t \le b\}$ has the same distribution as $\{\int K([(b-a)t+a]/h - s)\, dW(s) : 0 \le t \le 1\}$. Apply Corollary A1 of Bickel and Rosenblatt (1973) to the stationary Gaussian process

$$\left\{ C_{K,1}^{-1/2} \int K\left((b-a)t + \frac{a}{h} - s\right)\, dW(s) : 0 \le t \le \frac{1}{h} \right\}.$$

Theorem 1 now follows from (A.4) and (A.5), Lemma 1, and the fact that $Y_{n,1}$ and $Y_{n,2}$ have the same joint laws.

References

Alonso-García, J., Mateo, M., Sen, B., Banerjee, M., Catelan, M., Minniti, D. and von Braun, K. (2012). Uncloaking globular clusters in the inner galaxy. *Astron. J.* **143**, article 70.

Andersen, R. S. and de Hoog, F. R. (1990). Abel integral equations. *in* M. Golberg, ed., *Numerical Solutions of Integral Equations*. Plenum, New York, pp. 373–410.

Antoniadis, A., Fan, J. and Gijbels, I. (2001). A wavelet method for unfolding sphere size distributions. *Canad. J. Statist.* **29**, 251–268.

Bickel, P. J. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* **1**, 1071–1095.

Birke, M., Bissantz, N. and Holzmann, H. (2010). Confidence bands for inverse regression models. *Inverse Problems* **26**, article 115020.

Bissantz, N. and Birke, M. (2009). Asymptotic normality and confidence intervals for inverse regression models with convolution-type operators. *J. Multivariate Anal.* **100**, 2364–2375.

Bissantz, N., Dümbgen, L., Holzmann, H. and Munk, A. (2007). Non-parametric confidence bands in deconvolution density estimation. *J. Roy. Statist. Soc. B* **69**, 483–506.

Bissantz, N. and Holzmann, H. (2008). Statistical inference for inverse problems. *Inverse Problems* **24**, article 034009.

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

Csörgő, M. and Révész, P. (1981). *Strong Approximations in Probability and Statistics*. Academic Press, New York.

Delaigle, A., Hall, P. and Jamshidi, F. (2015). Confidence bands in nonparametric errors-in-variables regression. *J. Roy. Statist. Soc. B* **77**, 149–169.

Dony, J. and Einmahl, U. (2006). Weighted uniform consistency of kernel density estimators with general bandwidth sequences. *Electron. J. Probab.* **11**, 844–859.

Dudek, A. and Szkutnik, Z. (2008). Minimax unfolding of the spheres size distribution from linear sections. *Statist. Sinica* **18**, 1063–1080.

Eubank, R. L. and Speckman, P. L. (1993). Confidence bands in nonparametric regression. *J. Amer. Statist. Assoc.* **88**, 1287–1301.

Giné, E. and Nickl, R. (2010). Confidence bands in density estimation. *Ann. Statist.* **38**, 1122–1170.

Giné, E. and Nickl, R. (2016). *Mathematical Foundations of Infinite-Dimensional Statistical Models*. Cambridge University Press, New York.

Golubev, G. K. and Levit, B. Y. (1998). Asymptotically efficient estimation in the Wicksell problem. *Ann. Statist.* **26**, 2407–2419.

Groeneboom, P. and Jongbloed, G. (1995). Isotonic estimation and rates of convergence in Wicksell's problem. *Ann. Statist.* **23**, 1518–1542.

Groeneboom, P. and Jongbloed, G. (2014). *Nonparametric Estimation under Shape Constraints*. Cambridge University Press, New York.

Hall, P. (1992). Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *Ann. Statist.* **20**, 675–694.

Hall, P. and Smith, R. L. (1988). The kernel method for unfolding sphere size distributions. *J. Comput. Phys.* **74**, 409–421.

Klebaner, F. C. (2005). *Introduction to Stochastic Calculus With Applications*. Imperial College Press, London.

Lounici, K. and Nickl, R. (2011). Global uniform risk bounds for wavelet deconvolution estimators. *Ann. Statist.* **39**, 201–231.

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

Ripley, B. D. (1981). *Spatial Statistics*. Wiley, New York.

Sen, B. and Woodroofe, M. (2012). Bootstrap confidence intervals for isotonic estimators in a stereological problem. *Bernoulli* **18**, 1249–1266.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

Taylor, C. C. (1983). A new method for unfolding sphere size distributions. *J. Microsc.* **132**, 57–66.

van Es, B. and Hoogendoorn, A. (1990). Kernel estimation in Wicksell's corpuscle problem. *Biometrika* **77**, 139–145.

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

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

E-mail: jwojdyla@agh.edu.pl

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

E-mail: szkutnik@agh.edu.pl

(Received December 2015; accepted November 2016)