Electronic Journal of Statistics
Vol. 12 (2018) 194–223
ISSN: 1935-7524
https://doi.org/10.1214/18-EJS1391

Asymptotic confidence bands in the Spektor-Lord-Willis problem via kernel estimation of intensity derivative

Bogdan Ćmiel

Institute of Mathematics, Polish Academy of Sciences
ul. św. Tomasza 30, 31-027 Kraków, Poland
e-mail: cmielbog@gmail.com

Zbigniew Szkutnik and Jakub Wojdyła

AGH University of Science and Technology, Faculty of Applied Mathematics,
al. A. Mickiewicza 30, 30-059 Krakow, Poland
e-mail: szkutnik@agh.edu.pl; jwojdyla@agh.edu.pl

Abstract: The stereological problem of unfolding the distribution of spheres radii from linear sections, known as the Spektor-Lord-Willis problem, is formulated as a Poisson inverse problem and an L^2-rate-minimax solution is constructed over some restricted Sobolev classes. The solution is a specialized kernel-type estimator with boundary correction. For the first time for this problem, non-parametric, asymptotic confidence bands for the unfolded function are constructed. Automatic bandwidth selection procedures based on empirical risk minimization are proposed. It is shown that a version of the Goldenshluger-Lepski procedure of bandwidth selection ensures adaptivity of the estimators to the unknown smoothness. The performance of the procedures is demonstrated in a Monte Carlo experiment.

MSC 2010 subject classifications: Primary 62G05, 62G20, 45Q05.
Keywords and phrases: Spektor-Lord-Willis problem, inverse problem, minimax risk, rate of convergence, derivative estimation, kernel estimator, confidence bands, adaptive estimator.

Received September 2016.

1. Introduction

Consider a population of balls with random radii, randomly placed in an opaque medium and, hence, not directly observable. A linear section through the medium, like drilling through a rock or a muscle biopsy, allows for measuring the radii (half-lengths) of the line segments that are intersections of the line probe with the balls. The ultimate goal is to use those measurements to unfold both the balls radii distribution and the 'density' of the balls in the medium. Early formulation of the problem dates back to Spektor [30] and Lord and Willis [23] and was related to some measurements in material sciences. A review of some heuristic algorithms traditionally used for solving the Spektor-Lord-Willis problem (henceforth, SLW problem) and related stereological problems is given

in [10, Ch. 10.4.3]. Although planar rather than linear sections through the medium are generally preferred in experimental setups, leading to the better known Wicksell's problem, there are also important practical applications of line probes (e.g., [2, 21]). Moreover, linear intercept measurements on polished metallographic sections may lead to the SLW problem, even if plane sections through the medium are taken [22, p. 117].

Following [33] and [16], we formulate the SLW problem as a Poisson inverse problem. Assume that the balls radii distribution is supported in [0, 1] and the balls centers form a homogeneous Poisson process in $\mathbb{R}^3$ with intensity λ (all density and intensity functions are taken w.r.t. respective Lebesgue measures). If the balls radii distribution has a density, say ρ, then the observed radii of the line segments form a Poisson process on [0, 1] with intensity function $ns(u)$, where n is the 'size of the experiment', related to the total length of the observed linear probe, and

$$s(u) = 2u \int_u^1 f(x)\, dx =: (Sf)(u), \tag{1}$$

with $f(x) = \lambda\rho(x)$, see [33]. The goal is to unfold f from observed linear sections radii and to study asymptotics with $n \to \infty$, which is a special form of a Poisson inverse problem. With $S : L^2([0,1], dx) \to L^2([0,1], du)$, S is a compact Hilbert-Schmidt operator, its inverse is unbounded and the SLW problem is ill-posed in the Hadamard sense.

Poisson inverse problems were studied in some generality in, e.g., [31, 32, 1]. Spectral type solutions for the SLW problem, minimax on Sobolev ellipsoids were constructed in [16] and [34] and B-spline sieved quasi-maximum likelihood estimation was studied in [33]. Adaptive wavelet solutions, rate-minimax over some Besov balls were found in [11, 12].

The main new contribution of the present paper is the construction of non-parametric, asymptotic confidence bands for the unfolded intensity function. Construction of confidence bands in direct problems of function estimation started in 1973 with the seminal paper by Bickel and Rosenblatt [3], who constructed confidence bands for density estimated from an i.i.d. sample, and continued in several further developments, as summarized, e.g., in [19, Ch. 5.1.3] and [6]. The latter paper was also the first step towards the construction of confidence bands in inverse problems and was followed in recent years by several similar works [7, 4, 24, 8, 13, 25]. Despite their practical importance, no confidence bands had been constructed, however, for densities in stereological inverse problems until only recently, when such bands were produced in the Wicksell's problem in [36].

The methodology developed in [3, 6, 36] works well for estimators of kernel type. Therefore, in Section 2 we propose a new kernel-type estimator in the SLW problem and study its theoretical properties. Rate minimaxity over some Sobolev classes and adaptivity issues are studied in Section 3, and asymptotic, non-parametric confidence bands are constructed in Section 4. We also report in Section 5 on a practical implementation, including automatic bandwidth selection based on the empirical risk minimization principle, and on results of an

extensive Monte Carlo study of the proposed procedures. Some final conclusions are formulated in Section 6. Proofs related to adaptivity are presented in Appendix A.

It should be stressed that, in contrast to all the previous constructions of confidence bands in inverse problems, we construct the bands in the Poisson inverse problem setup with random number of data points, which is more realistic in stereology than inverse density estimation setup with non-random sample size.

2. The estimator

Our first goal is to construct an L^2-rate-minimax estimator of the intensity function f, given observations of a Poisson process with the intensity function ns, related to f via (1). The main idea of our approach is based on the observation that this problem may be reformulated as the problem of derivative estimation with a weighted L^2-loss function. Let $X_1, \ldots, X_{N(n)}$ be random points associated with the Poisson process with intensity ns. Then $X_1^2, \ldots, X_{N(n)}^2$ form a Poisson process with intensity ng, where $g(u) = s(\sqrt{u})/(2\sqrt{u})$ and

$$g(u) = \int_{\sqrt{u}}^1 f(x)\,dx =: (Gf)(u).$$

For continuous f, $g'(u) = -f(\sqrt{u})/(2\sqrt{u})$ and $f(x) = -2xg'(x^2)$. Hence, any estimator $\hat{g}'$ of g' gives a natural estimator $\hat{f}(x) := -2x\hat{g}'(x^2)$ of f and

$$\int_0^1 \left[\hat{f}(x) - f(x)\right]^2 dx = \int_0^1 \left[\hat{g}'(u) - g'(u)\right]^2 2\sqrt{u}\,du.$$

To express the smoothness assumptions, the functions f and g will be further considered as functions defined on $\mathbb{R}$. The L^2 and L^∞ norms will be denoted with $\|\cdot\|_2$ and $\|\cdot\|_\infty$. For some $m \in \mathbb{N}$, define a Sobolev-type class of functions

$$W(m, L) = \left\{ s : \mathbb{R} \to \mathbb{R} : s^{(m-1)} \text{ is absolutely continuous and } ||s^{(m)}||_2 \leqslant L \right\}.$$

To evaluate the estimation risk for f in $W(m, L)$ with standard methods, as in, e.g., [35], it would be convenient to have g in $W(m+1, L)$. However, since $g(0) = \int_0^1 f(x)\,dx$, any g is necessarily discontinuous at zero. To circumvent the problem, we will use the reflection device (e.g., [29, Ch. 2.10]) and estimate the function g' under the assumption that the symmetrized g, i.e. the function $u \to g_{sym}(u) := g(|u|)$, belongs to $W(m+1, L)$. Define

$$\mathcal{F}_{m,C} = \{f : \mathbb{R} \to [0, \infty) : \operatorname{supp} f \subset [0, 1] \text{ and } (Gf)_{sym} \in W(m+1, C)\}.$$

It is elementary to see that, if f has a left limit $f(1^-)$ at 1, then the left derivative $g'(1^-)$ of g at 1 exists and $g'(1^-) = -f(1^-)/2$. Because $m \geq 1$, g' must be continuous on $\mathbb{R}$ for $f \in \mathcal{F}_{m,C}$, which implies that $f(1^-) = 0$. Further,

it follows by induction that, for $k = 1, \ldots, m+1$, there exist absolute constants $C_0^k, \ldots, C_{k-1}^k$ such that, for almost every $u > 0$ and every $f \in W(m, L)$,

$$g^{(k)}(u) = \sum_{i=0}^{k-1} C_i^k \, \frac{f^{(k-1-i)}(\sqrt{u})}{\sqrt{u^{k+i}}}, \tag{2}$$

so that the class $\mathcal{F}_{m,C}$ contains the functions $f \in W(m, L)$ with derivatives tending to zero sufficiently fast, when the argument goes to zero.

Proposition **1.** *Let $f \in W(m, L)$, $\operatorname{supp} f \subset [0, 1]$ and $f \geqslant 0$. Assume that there exist $\varepsilon > 0$, $a > 0$ and $\xi > 0$ such that*

$$|f^{(m)}(x)| \leqslant a x^{m+\xi}, \quad \text{for} \quad x \in [0, \varepsilon).$$

Then, $f \in \mathcal{F}_{m,C}$ for some C dependent on $m, L, \varepsilon, a, \xi$.

Proof. Since $f^{(i)}(0) = 0$ for $i = 0, \ldots, m-1$, one has, for every $u \in [0, \varepsilon)$, $k = 0, \ldots, m-1$, and with some $\gamma \in [0, 1]$,

$$\begin{aligned} \left| f^{(k)}(u) \right| &= \left| \sum_{i=0}^{m-1-k} f^{(i+k)}(0) \frac{u^i}{i!} + f^{(m)}(\gamma u) \frac{u^{m-k}}{(m-k)!} \right| \leqslant a \, [\gamma u]^{m+\xi} \frac{u^{m-k}}{(m-k)!} \\ &\leqslant B_k(a, m) \, u^{2m-k+\xi}. \end{aligned}$$

Using this inequality in equation (2), we obtain

$$\left| g^{(m+1)}(u) \right| \leqslant \sum_{i=0}^{m} \left| C_i^{m+1} \right| B_{m-i}(a, m) \sqrt{u^{\,\xi-1}} = A_{m+1}(a) \sqrt{u^{\,\xi-1}}$$

and

$$\left| g^{(m)}(u) \right| \leqslant A_m(a) \sqrt{u^{\xi+1}},$$

for every $u \in (0, \varepsilon^2)$. Since $f^{(k)}$ is absolutely continuous for $k = 0, \ldots, m-1$, it is obvious that $g^{(m)}$ is absolutely continuous. We also have

$$\begin{aligned} \int_0^1 \left[g^{(m+1)}(u) \right]^2 du &\leqslant \int_0^{\varepsilon^2} A_{m+1}^2(a) u^{\,\xi-1} \, du \; + \int_{\varepsilon^2}^1 \left[g^{(m+1)}(u) \right]^2 du \\ &\leqslant A_{m+1}^2(a) \frac{\varepsilon^{2\xi}}{\xi} + (m+1) \sum_{i=0}^{m} (C_i^{m+1})^2 \int_{\varepsilon^2}^1 \frac{\left[f^{(m-i)}(\sqrt{u}) \right]^2}{u^{m+1+i}} \, du \\ &\leqslant A_{m+1}^2(a) \frac{\varepsilon^{2\xi}}{\xi} + 2(m+1) \sum_{i=0}^{m} \frac{(C_i^{m+1})^2}{\varepsilon^{2m+2i+1}} \int_{\varepsilon}^1 \left[f^{(m-i)}(x) \right]^2 dx \\ &\leqslant C(m, L, \varepsilon, a, \xi), \end{aligned}$$

which completes the proof. □

Following the idea sketched above, we start with a kernel estimator of the derivative g', based on symmetrized sample $\pm X_1^2, \ldots, \pm X_{N(n)}^2$, and obtain the estimator for f in the form

$$\hat{f}_n(x) = -\frac{2x}{nh^2} \sum_{i=1}^{N(n)} \left[K'\left(\frac{x^2 - X_i^2}{h}\right) + K'\left(\frac{x^2 + X_i^2}{h}\right) \right] \mathbb{1}_{[0,1]}(x), \tag{3}$$

with some kernel K and some bandwidth h. For the minimaxity of the estimator, it is convenient to make h a function of the non-random experiment size n. For the construction of confidence bands in Section 4, however, it will be more natural to make h dependent on the random number $N(n)$ of data points. As $N(n)/n$ tends a.s. to a limit (depending on f), those two types of assumptions are closely related.

Throughout this article, we define the order of a kernel as the order of its first non-zero moment. At several places, we impose

Assumption **1.** *K is an absolutely continuous, symmetric kernel with square-integrable derivative K' and such that $\int K(u)\,du = 1$.*

Theorem **1.** *If K is a kernel of order at least m, such that $\int |u^m K(u)|\,du < \infty$ and Assumption 1 is fulfilled, and if $h \asymp n^{-1/(2m+3)}$, then there exists a constant $D_{m,C}$ such that for every $f \in \mathcal{F}_{m,C}$*

$$\mathbb{E}||\hat{f}_n - f||_2^2 \leqslant D_{m,C}\, n^{\frac{-2m}{2m+3}}.$$

Proof. Let $p(u) := g(|u|)$ and let

$$\hat{p'}(u) := \frac{1}{nh^2} \sum_{i=1}^{N(n)} \left[K'\left(\frac{u - X_i^2}{h}\right) + K'\left(\frac{u + X_i^2}{h}\right) \right]$$

be the standard kernel estimator of p', based on the symmetrized sample. Define $\hat{g'}(u) = \hat{p'}(u)\mathbb{1}_{[0,1]}(u)$. Then

$$\begin{aligned} \int_0^1 \left[\hat{g'}(u) - g'(u)\right]^2 2\sqrt{u}\,du &\leqslant 2\int_0^1 \left[\hat{g'}(u) - g'(u)\right]^2 du \\ &\leqslant \int_{\mathbb{R}} \left[\hat{p'}(u) - p'(u)\right]^2 du, \end{aligned}$$

and, consequently,

$$\mathbb{E}||\hat{f}_n - f||_2^2 \leqslant \mathbb{E}||\hat{p'} - p'||_2^2.$$

With $b(x) = \mathbb{E}\hat{p'}(x) - p'(x)$ and $\sigma^2(x) = \mathbb{E}[\hat{p'}(x) - \mathbb{E}\hat{p'}(x)]^2$, one has the standard variance-bias decomposition of the risk

$$\mathbb{E}||\hat{p'} - p'||_2^2 = \int_{\mathbb{R}} \sigma^2(x)\,dx + \int_{\mathbb{R}} b^2(x)\,dx.$$

Let

$$\begin{aligned}\eta_i(x) &= \left[K'\left(\frac{x-X_i^2}{h}\right)+K'\left(\frac{x+X_i^2}{h}\right)\right] \\ &\quad -\mathbb{E}\left[K'\left(\frac{x-X_i^2}{h}\right)+K'\left(\frac{x+X_i^2}{h}\right)\right],\end{aligned}$$

for $i=1,\ldots,N(n)$. Then, representing the Poisson process as a mixed empirical process (cf. [26, Th. 1.2.1]),

$$\begin{aligned}\sigma^2(x) &= \mathbb{E}\left(\frac{1}{nh^2}\sum_{i=1}^{N(n)}\eta_i(x)\right)^2 = \mathbb{E}\left\{\mathbb{E}\left[\left(\frac{1}{nh^2}\sum_{i=1}^{N(n)}\eta_i(x)\right)^2\Big|N(n)\right]\right\} \\ &= \frac{1}{n^2h^4}\mathbb{E}\left[N(n)\mathbb{E}\left(\eta_1^2(x)|N(n)\right)\right] \leqslant \frac{c}{nh^4}\mathbb{E}\left[\eta_1^2(x)\right],\end{aligned}$$

where $\mathbb{E}[N(n)]/n = c = \int_0^1 g(x)\,dx$, which is bounded because $p=(Gf)_{sym}\in W(m+1,C)$. Since

$$\begin{aligned}\mathbb{E}\left[\eta_1^2(x)\right] &\leqslant \mathbb{E}\left[K'\left(\frac{x-X_1^2}{h}\right)+K'\left(\frac{x+X_1^2}{h}\right)\right]^2 \\ &\leqslant 2\int_{\mathbb{R}}\left[K'\left(\frac{x-z}{h}\right)\right]^2\frac{g(z)}{c}\,dz + 2\int_{\mathbb{R}}\left[K'\left(\frac{x+z}{h}\right)\right]^2\frac{g(z)}{c}\,dz,\end{aligned}$$

we have

$$\begin{aligned}\int_{\mathbb{R}}\sigma^2(x)\,dx &\leqslant \frac{2c}{nh^4}\int_{\mathbb{R}}\left\{\int_{\mathbb{R}}\left[K'\left(\frac{x-z}{h}\right)\right]^2dx\right. \\ &\qquad \left.+\int_{\mathbb{R}}\left[K'\left(\frac{x+z}{h}\right)\right]^2dx\right\}\frac{g(z)}{c}\,dz \\ &= \frac{4c}{nh^3}\int_{\mathbb{R}}\left[K'(u)\right]^2du = \frac{C_1(c)}{nh^3}.\end{aligned}$$

Now, we evaluate the bias, using the absolute continuity of K in the integration by parts step:

$$\begin{aligned}b(x) &= \frac{1}{nh^2}\mathbb{E}\left\{N(n)\mathbb{E}\left[K'\left(\frac{x-X_1^2}{h}\right)+K'\left(\frac{x+X_1^2}{h}\right)\Big|N(n)\right]\right\}-p'(x) \\ &= \frac{c}{h^2}\left[\int_0^1 K'\left(\frac{x-z}{h}\right)\frac{g(z)}{c}\,dz+\int_{-1}^0 K'\left(\frac{x-z}{h}\right)\frac{g(-z)}{c}\,dz\right]-p'(x) \\ &= \frac{1}{h^2}\int_{\mathbb{R}}K'\left(\frac{x-z}{h}\right)p(z)\,dz-p'(x) = \frac{1}{h}\int_{\mathbb{R}}K\left(\frac{x-z}{h}\right)p'(z)\,dz-p'(x) \\ &= \int_{\mathbb{R}}K(u)\left[p'(x-uh)-p'(x)\right]du.\end{aligned}$$

Standard calculation with $p \in W(m+1, C)$, analogous to that on page 14 in [35], gives after Taylor expansion of $p'(x-uh) - p'(x)$ and double application of the generalized Minkowski inequality

$$\int_{\mathbb{R}} b^2(x)\, dx \leq C_2(m, C) h^{2m}.$$

With $h \asymp n^{\frac{-1}{2m+3}}$, this finally gives

$$\mathbb{E}||\hat{f}_n - f||_2^2 \leqslant \frac{C_1(C)}{nh^3} + C_2(m, C) h^{2m} \leqslant D_{m,C}\; n^{\frac{-2m}{2m+3}}. \qquad \square$$

3. Lower bounds, minimaxity and adaptivity

For a good lower bound, a possibly large number of well separated functions in $\mathcal{F}_{m,C}$ should be constructed, for which the corresponding data distributions are close to each other. As in [11], the construction will be based on smooth wavelets. Let $\psi \in C^m$ be a compactly supported mother wavelet, e.g. a Daubechies wavelet, and let $\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k)$, with $j, k \in \mathbb{Z}$. Suppose that $\operatorname{supp} \psi = [0, S(m)]$ for some $S(m) \in \mathbb{N}$. For some $L > 0$, define

$$\mathcal{G}_{m,L}(j) := \left\{ f_\omega : f_\omega = f_0 + \delta_j \sum_{k \in \mathcal{G}_j(\psi)} \omega_k \psi_{jk} \right\},$$

where

1. $f_0 \in W(m, L/2)$ is a fixed, non-negative function supported on $[1/8, 1]$ and such that $\min_{x \in [1/4, 1/2]} f_0(x) = C_3 > 0$,
2. $\int_{1/2}^1 f_0(x)\, dx = C_4 > 0$,
3. $\omega = (\ldots, \omega_{-1}, \omega_0, \omega_1, \ldots)$ with $\omega_k \in \{0, 1\}$, for $k \in \mathbb{Z}$,
4. $\mathcal{G}_j(\psi) := \left\{ k \in \mathbb{Z} : \operatorname{supp} \psi_{jk} \subset \left[\frac{1}{4}, \frac{1}{2}\right] \right\}$,
5. $\delta_j = \min\left\{ \frac{L}{S(m)||\psi^{(m)}||_\infty} 2^{-j(m+\frac{1}{2})}, \frac{C_3}{S(m)||\psi||_\infty} 2^{-j/2} \right\}$.

Note that, for every j and x, the cardinality of the set $\{k \in \mathbb{Z} : \psi_{jk}^{(m)}(x) \neq 0\}$ does not exceed $S(m)$ and $\psi_{jk}^{(m)}(x) = 2^{j(m+\frac{1}{2})}\psi^{(m)}(2^j x - k)$. It is now easy to see that $\mathcal{G}_{m,L}(j) \subset \mathcal{F}_{m,C}$ for all $j \in \mathbb{Z}$ and with some $C > 0$. Indeed, for all $f_\omega \in \mathcal{G}_{m,L}(j)$, we have

$$\begin{aligned} ||f_\omega^{(m)}||_2 &\leqslant ||f_0^{(m)}||_2 + \delta_j \sum_{k \in \mathcal{G}_j(\psi)} \|\psi_{jk}^{(m)}\|_2 \\ &\leqslant \frac{L}{2} + \frac{L}{S(m)||\psi^{(m)}||_\infty} \sum_{k \in \mathcal{G}_j(\psi)} \left(\int_{1/4}^{1/2} \left| \psi^{(m)}(2^j x - k) \right|^2 dx \right)^{1/2} \end{aligned}$$

$$\leqslant \quad \frac{L}{2} + \frac{L}{2} = \; L$$

and it is obvious that each f_ω is non-negative because of the condition 5 above and because, again, for every fixed j and x, the cardinality of the set $\{k \in \mathbb{Z} : \psi_{jk}(x) \neq 0\}$ does not exceed $S(m)$. Since supp $f_\omega \subset [1/8, 1]$, using Proposition 1, we conclude that $f_\omega \in \mathcal{F}_{m,C}$. Reasoning further as in the proof of Proposition 1 in [11], with $2^j \asymp n^{1/(2m+3)}$ and using conditions 2 and 5, we obtain, via an application of the Assouad's cube technique, the existence of a constant $C_5 > 0$ such that

$$\sup_{f \in \mathcal{F}_{m,C}} \mathbb{E}||\tilde{f}_n - f||_2^2 \geqslant C_5 \; n^{\frac{-2m}{2m+3}}$$

for any estimator $\tilde{f}_n$. The constant C may influence the value of C_5, but not the rate. This means that the estimator $\hat{f}_n$ in Theorem 1 is rate-minimax over $\mathcal{F}_{m,C}$. Moreover, this also implies that, for any estimator $\tilde{f}_n$, $\sup_{f \in \mathcal{F}_{m,C}} \mathbb{E}\|\tilde{f}_n - f\|_\infty$ cannot approach zero faster than $n^{-m/(2m+3)}$, which is more relevant for the discussion of uniform confidence bands. It will be seen in the next section that the width of our confidence bands converges to zero at the rate $n^{-(m/(2m+3)-\xi)}$, with ξ that may be arbitrarily small, but positive because of undersmoothing used for bias control.

The bandwidth depends on the generally unknown smoothness of the estimated function, but the estimator can be made adaptive through a suitable data-driven choice of the bandwidth. A natural choice for adaptive bandwidth selection for kernel estimators is the Goldenshluger-Lepski method (cf. [20]). For brevity, call such adaptive estimators 'GL-estimators'. A version of GL-estimators has been studied for probability densities and their derivatives in the i.i.d. setup in [15]. For Poisson process intensity functions (but not for their derivatives), GL-estimators have been studied in [28]. In order to produce an oracle inequality for our boundary corrected estimator of the derivative of the intensity of the observable Poisson process (and, consequently, for the unfolded intensity), we adapt to the Poisson process setup with boundary correction the proof of Proposition 3 in [15], pretty much in the same spirit, in which Proposition 2 in [15] has been ported to the Poisson process setup as Theorem 2 in [28].

As in the proof of Theorem 1, we work with symmetrized functions p, p' and with the estimator $\hat{p'}_h$, with the index that marks the dependence on the bandwidth. Similarly, we write in this section $\hat{f}_h$ rather than $\hat{f}_n$. For the reflection device, it is natural to assume K symmetric. This makes $\hat{p'}_h$ antisymmetric and simplifies some technical arguments in the proofs.

From the proof of Theorem 1,

$$\mathbb{E}\hat{p'}_h = K_h \star p' \quad \text{and} \quad \mathbb{E}\|\hat{p'}_h - K_h \star p'\|_2^2 = \int_{\mathbb{R}} \mathrm{V}ar\, \hat{p'}_h \, dx \leq \frac{4\|g\|_1 \|K'\|_2^2}{nh^3},$$

where $K_h(\cdot) = h^{-1}K(\cdot/h)$, $\star$ denotes the convolution and $\|\cdot\|_1$ stands for the L^1-norm. For a function, say v, denote with $\|v\|_*$ the L^2-norm with respect to

$2\sqrt{u}du$ of v truncated to the $[0,1]$ domain. The ultimate goal is to approximate h that minimizes $\mathbb{E}\|\hat{f}_h - f\|_2^2$. This will be achieved through minimization of some tight upper bound, say $\Phi(h)$, for the lower bound $\mathbb{E}\|\hat{f}_h - f\|_2$ to $(\mathbb{E}\|\hat{f}_h - f\|_2^2)^{1/2}$, with suitably estimated unknown quantities. By triangle and Cauchy-Schwarz inequalities, and because $\|\cdot\|_* \le \|\cdot\|_2$ for antisymmetric functions,

$$\mathbb{E}\|\hat{f}_h - f\|_2 = \mathbb{E}\|\hat{p'}_h - p'\|_* \le \|K_h \star p' - p'\|_* + \frac{2\|g\|_1^{1/2}\|K'\|_2}{\sqrt{nh^3}} =: \Phi(h).$$

The oracle bandwidth is defined as the minimizer of $\Phi(h)$ over a set, say $\mathcal{H}$, of candidate bandwidths. It depends on the unknown p. The idea of the GL method is to minimize a specific estimator of the function $\Phi(h)$. Set

$$\hat{p}_h(x) = \frac{1}{n}\sum_{i=1}^{N(n)} \left[K_h(x - X_i^2) + K_h(x + X_i^2)\right],$$

the kernel estimator of p. Then $\hat{p'}_h = (\hat{p}_h)'$ and, since $K_h \star p' = (K_h \star p)'$, a natural estimator of this term may be defined as

$$\begin{aligned}\hat{p'}_{h,h'}(x) &:= (K_h \star \hat{p}_{h'})'(x)\\ &= \frac{1}{n}\sum_{i=1}^{N(n)} \left[(K_h \star K_{h'})'(x - X_i^2) + (K_h \star K_{h'})'(x + X_i^2)\right],\end{aligned}$$

with some fixed h'. Further, since $\mathbb{E}N(n) = n\|g\|_1$, a natural estimator of $\|g\|_1$ is $N(n)/n$. In the GL method, the secondary bandwidth h' is eliminated by minimizing w.r.t. h an 'upper envelope' of a family of functions indexed by h' and suitably constructed to properly estimate the bias term.

With some $\eta > 0$, define $\chi := (1+\eta)(1+\|K\|_1)\|K'\|_2$ and

$$A(h) := \sup_{h' \in \mathcal{H}} \left\{\|\hat{p'}_{h,h'} - \hat{p'}_{h'}\|_* - \frac{\chi\sqrt{N(n)}}{n(h')^{3/2}}\right\}_+,$$

and set

$$\hat{h} := \arg\min_{h \in \mathcal{H}} \left\{A(h) + \frac{\chi\sqrt{N(n)}}{nh^{3/2}}\right\}.$$

For the GL estimator of p', defined as $\hat{p'}_{GL} := \hat{p'}_{\hat{h}}$, we have the following oracle inequality proved in Appendix A.

Proposition **2.** *Assume that K satisfies Assumption 1. If $\mathcal{H} = \{D^{-1} : D = 1, \ldots, \lfloor(\delta n)^{1/2}\rfloor\}$, with some $\delta > 0$, then, for all n,*

$$\mathbb{E}\|\hat{p'}_{GL} - p'\|_*^2 \le C_3 \inf_{h \in \mathcal{H}} \left\{\|K_h \star p' - p'\|_2^2 + \frac{4\|g\|_1\|K'\|_2^2}{nh^3}\right\} + \frac{C_4}{n},$$

where $C_3 = 1 + 12(1+\eta)^2(1+\|K\|_1)^2$ and C_4 is a constant that depends only on the kernel and on $\|g\|_\infty$ and $\|g\|_1$.

This easily gives the adaptivity of the GL estimator of f, naturaly defined as

$$\hat{f}_{GL}(x) := -2x\hat{p'}_{GL}(x^2)\mathbb{1}_{[0,1]}(x).$$

THEOREM **2.** *If K is a kernel of order at least m such that $\int |u^m K(u)|\, du < \infty$, Assumption 1 is fulfilled and $\mathcal{H} = \{D^{-1} : D = D_{min}, \dots, D_{max}\}$, with $1 \le D_{min} \le \lfloor n^{1/(2m+3)} \rfloor$ and $D_{max} = \lfloor n^{1/2} \rfloor$, then for all $s = 1, \dots, m$,*

$$\mathbb{E}\|\hat{f}_{GL} - f\|_2^2 \asymp n^{-\frac{2s}{2s+3}}, \tag{4}$$

if $f \in \mathcal{F}_{s,C}$.

Note that, for $f \in \mathcal{F}_{s,C}$, the bias term $\|K_h \star p' - p'\|_2^2$ can be bounded, as in the proof of Theorem 1, by $C_2(s, C)h^{2s}$ and the leading constants in (4) depend on f only through $\|g\|_1$. Since, in the SLW problem, $\|g\|_1 \le \|g\|_\infty = \|f\|_1$, the rates are uniform over intersections of $\mathcal{F}_{s,C}$ with any fixed ball in L^1. Note also that conditions of Proposition 1 ensure that $\|f\|_1$ is upper bounded by a constant that depends on L only.

4. Confidence bands

The goal of this section is to construct asymptotic confidence bands for the intensity function f on an interval $[a, b] \subset [0, 1]$, with $a > 0$ and $b < 1$, based on the data $X_1^2, \dots, X_{N(n)}^2$.

If $f_0(x) = f(\sqrt{x})/(2\sqrt{x})$ corresponds to the intensity function of the Poisson process of the unobserved squared spheres radii, then, for continuous f_0, $f_0(u) = -g'(u)$. Therefore, to adapt the idea of Bickel and Rosenblatt ([3]), who proposed a method of construction of confidence bands around kernel estimators of a specific type, it is convenient to first construct confidence bands for f_0 on $[a^2, b^2]$ around a suitable kernel estimator of the derivative $-g'$, and then transform the results back to the original problem.

More specifically, let

$$\hat{f}_{0,n}(x) = -\frac{1}{nh^2} \sum_{i=1}^{N(n)} \left[K'\left(\frac{x - X_i^2}{h}\right) + K'\left(\frac{x + X_i^2}{h}\right) \right] \mathbb{1}_{[0,1]}(x)$$

be an estimator of f_0, similar to the estimator $\hat{f}_n$ of f defined in (3), but with a random bandwidth $h = h\,[N(n)]$ that depends on the random sample size $N(n)$. The first step is to investigate the asymptotic distribution of the maximal deviation of the estimator $\hat{f}_{0,n}$ from its mean over $[a^2, b^2]$. Define the process

$$Z_{N(n)}(t) = -\frac{nh^{3/2}}{N(n)^{1/2}\tilde{q}_{N(n)}(t)^{1/2}} \left[\hat{f}_{0,n}(t) - \mathbb{E}\hat{f}_{0,n}(t) \right], \quad t \in [a^2, b^2],$$

where $h = h\,[N(n)]$ and $\tilde{q}_{N(n)}$ is an appropriate estimator of the density $q(x) = g(x)/c$ of the observations X_i^2 (see Assumption 2(c) below).

The main results of this section depend on two sets of assumptions. The first and less restrictive one allows for construction of confidence bands for $\mathbb{E}\hat{f}_{0,n}$. The second set includes stronger assumptions needed for bias-correction and for construction of confidence bands for f itself. Denote with $\|\cdot\|_{\infty,I}$ the sup-norm on the interval $I = [a^2, b^2]$.

Assumption **2.**

(a) K is supported and twice continuously differentiable on $[-1, 1]$,

(b) h is a function such that $h(k) \to 0$ and $kh(k)^2/\log\{1/h(k)\} \to \infty$ as $k \to \infty$,

(c) the estimator $\tilde{q}_k$ of q satisfies $\|\tilde{q}_k - q\|_{\infty,I} = o_p\left(1/\log\{1/h(k)\}\right)$ as the sample size $k \to \infty$, with h as in assumption (b),

(d) the intensity function f is continuous on $(a - \Delta, b + \Delta)$, for some $\Delta > 0$, and $\int_b^1 f(x)\,dx > 0$.

Assumption **3.**

(a) For some integer $m \geq 1$, K is a kernel of order at least m, supported and twice continuously differentiable on $[-1, 1]$,

(b) the bandwidth h satisfies $h \asymp N(n)^{-\gamma}$, with $\gamma \in (\frac{1}{2m+3}, \frac{1}{2})$,

(c), (d) same as Assumptions 2(c) and 2(d), respectively,

(e) for m as in assumption (a) and for some $\Delta > 0$, f is $(m-1)$-times continuously differentiable in $(a-\Delta, b+\Delta)$ and there exists bounded $f^{(m)}$ in $(a - \Delta, b + \Delta)$.

With $h(k) \asymp k^{-\gamma}$, Assumption 2(c) reduces to $\|\tilde{q}_k - q\|_{\infty,I} = o_p\left(1/\log k\right)$, which is typically satisfied by, e.g., kernel estimators with polynomial rate of convergence (cf. [14]).

For any $a > 0$ and $b < 1$, the conditions imposed on the intensity function f hold for most commonly assumed distributions, including SML-A (with $m = 1$) and SML-B, NM, Beta(4,2) (with any $m \geq 1$), used in our simulation studies in Section 5.

Note that, due to Assumption 2(d), g is bounded away from zero on $[a^2, b^2]$, because it is non-increasing and $g(b^2) = \int_b^1 f(x)\,dx$, which also implies that $g^{1/2}$ has bounded derivative in $[a^2 - \Delta, b^2 + \Delta]$, with some $\Delta > 0$. Also, due to Assumption 3(e), g is m-times continuously differentiable in $[a^2 - \Delta, b^2 + \Delta]$ and there exists bounded $g^{(m+1)}$ in $(a^2 - \Delta, b^2 + \Delta)$. All those features of g are used in the proofs of Theorem 3 and Corollary 1 below.

The limiting distribution of the supremum of the process $\{Z_{N(n)}(t) : a^2 \leq t \leq b^2\}$ is given in the following theorem.

Theorem **3.** *Under Assumption 2, for each $x \in \mathbb{R}$,*

$$P\left([2\log(1/h)]^{1/2}\left[\|Z_{N(n)}\|_{\infty,I}/C_{K,1}^{1/2} - d_{N(n)}\right] \leq x\right) \to \exp\left[-2\exp(-x)\right]$$

as $n \to \infty$, where

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

and

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

Proof. The proof will only be sketched briefly, because large parts of it are similar to the proofs in [36]. The first part of the proof goes, conditionally on $N(n) = k$, along the same lines as the proof of Theorem 1 in [36] with the estimator

$$\tilde{f}_{0,k}(x) = -\frac{1}{h^2} \sum_{i=1}^{k} \left[K'\left(\frac{x - X_i^2}{h}\right) + K'\left(\frac{x + X_i^2}{h}\right) \right] \mathbb{1}_{[0,1]}(x), \quad h = h(k),$$

based on i.i.d. sample $X_1, \ldots, X_k$ and with the process

$$\begin{aligned} Y_k(t) &= -\frac{h^{3/2}}{k^{1/2} q(t)^{1/2}} \left[\tilde{f}_{0,k}(t) - \mathbb{E}\tilde{f}_{0,k}(t) \right] \\ &= \frac{h^{-1/2}}{q(t)^{1/2}} \int_0^1 \left[K'\left(\frac{t-x}{h}\right) + K'\left(\frac{t+x}{h}\right) \right] d\alpha_k(x), \quad t \in [a^2, b^2], \end{aligned}$$

where α_k is the empirical process corresponding to the distribution function of $X_1^2, \ldots, X_k^2$. After noting that asymptotically, for $h \le a^2$,

$$Y_k(t) = \frac{h^{-1/2}}{q(t)^{1/2}} \int_0^1 K'\left(\frac{t-x}{h}\right) d\alpha_k(x), \quad t \in [a^2, b^2],$$

one can construct appropriate approximations $Y_{k,0}$, $Y_{k,1}$, $Y_{k,2}$, and $Y_{k,3}$ of the process Y_k, and obtain, for each $x \in \mathbb{R}$,

$$P\left([2\log(1/h)]^{1/2} \left[\|Y_k\|_{\infty,I} / C_{K,1}^{1/2} - d_k \right] \le x \right) \to \exp\left[-2\exp(-x)\right]$$

as $k \to \infty$, where

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

The unknown quantity in Y_k is $q(t)$, with $t \in [a^2, b^2]$. However, reasoning as in the proof of Corollary 1 in [36], one deduces, using Assumption 2(c), that the above result remains true after replacing Y_k with

$$Y_{k,4}(t) = -\frac{h^{3/2}}{k^{1/2} \tilde{q}_k(t)^{1/2}} \left[\tilde{f}_{0,k}(t) - \mathbb{E}\tilde{f}_{0,k}(t) \right], \quad t \in [a^2, b^2].$$

Now, it follows from Lemma 4.1.1 in [26] that the same limiting distribution is valid for the process

$$T(N_k) = [2\log(1/h)]^{1/2} \left[\|Y_{k,4}\|_{\infty,I} / C_{K,1}^{1/2} - d_k \right], \qquad h = h(k),$$

with $k = N(n)$. An obvious modification of the results obtained for the estimator $\tilde{f}_{0,k}$, with $k = N(n)$, gives the required result for the estimator $\hat{f}_{0,n}$ and completes the proof. $\square$

To finish the construction of the confidence bands for f, one has to deal with the bias of $\hat{f}_{0,n}$, and then transform the obtained bands for f_0 to bands for f. To control the impact of the bias, we follow one of the major strategies which is undersmoothing: accepting less smooth estimator in order to reduce the bias (cf., e.g., [6], [5], [4], [25], [19, Ch. 6.4.2], [36]). (An alternative approach is explicit bias correction, see, e.g. [17].) In our case, this idea is realized in Assumption 3(b), which becomes clear after comparison with the rate of convergence of h to zero imposed in Theorem 1.

Summarizing, the confidence bands constructed using an undersmoothing bandwidth have the form implied by the following corollary.

Corollary **1.** *Under Assumption 3, for each $x \in \mathbb{R}$,*

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

as $n \to \infty$, where

$$l_n(t,x) = \frac{2t\tilde{q}_{N(n)}(t^2)^{1/2}N(n)^{1/2}C_{K,1}^{1/2}}{nh^{3/2}}\left[\frac{x}{[2\log(1/h)]^{1/2}} + d_{N(n)}\right],$$

the estimator $\hat{f}_n$ is defined by Equation 3, with $h \asymp N(n)^{-\gamma}$, and the constants $C_{K,1}$ and $d_{N(n)}$ are defined in Theorem 3.

Proof. For $x \in [a^2, b^2]$,

$$\mathbb{E}\hat{f}_{0,n}(x) = \mathbb{E}\left[\mathbb{E}\left(\hat{f}_{0,n}(x) \mid N(n)\right)\right] \\ = -\mathbb{E}\left[\frac{N(n)}{cn}\left(\int_{(x-1)/h}^{x/h} K(s)g'(x-hs)\,ds + \int_{x/h}^{(x+1)/h} K(s)g'(hs-x)\,ds\right)\right]$$

and, because $\operatorname{supp} K = [-1,1]$, for $x \in [a^2, b^2]$ and sufficiently small h one obtains

$$\mathbb{E}\hat{f}_{0,n}(x) = -\mathbb{E}\left(\frac{N(n)}{cn}\int_{-1}^{1} K(s)g'(x-hs)\,ds\right).$$

Therefore, standard reasoning based on the Taylor expansion of $g'(x-hs)$ around x shows that, under Assumptions 3(a), (b), (d), and (e), the bias of $\hat{f}_{0,n}$ is bounded by

$$|\mathbb{E}\hat{f}_{0,n}(x) - f_0(x)| = O(n^{-m\gamma}),$$

uniformly in $x \in [a^2, b^2]$.

Consequently, the difference between

$$T_{n,1} = [2\log(1/h)]^{1/2}\left[\|Z_{N(n)}\|_{\infty,I}/C_{K,1}^{1/2} - d_{N(n)}\right]$$

and

$$T_{n,2} = [2\log(1/h)]^{1/2}\left[\frac{nh^{3/2}}{N(n)^{1/2}\|\tilde{q}_{N(n)}\|_{\infty,I}^{1/2}}\|\hat{f}_{0,n} - f_0\|_{\infty,I}/C_{K,1}^{1/2} - d_{N(n)}\right]$$

is of the order of $O_p\left[n^{-(m+3/2)\gamma+1/2}(\log n)^{1/2}\right]$ and, for $\gamma > 1/(2m+3)$ as in Assumption 3(b), converges to zero in probability.

Finally, it follows from Theorem 3, by Slutsky's theorem, that, for each $x \in \mathbb{R}$,

$$P\left(T_{n,2} \leq x\right) \to \exp\left[-2\exp(-x)\right],$$

as $n \to \infty$, and Corollary 1 now follows by rearranging the terms and the substitution $f_0(x) = f(\sqrt{x})/(2\sqrt{x})$. □

5. Bandwidth selection and simulations

The Goldenshluger-Lepski (GL) procedure, although attractive theoretically, proved to be computationally very expensive. Even with effective, parallel implementation on a large grid of processors, it was hardly feasible to conduct Monte Carlo experiments of conclusive size, and this only for Gaussian kernels, for which convolutions can be computed analytically in closed forms. The necessity of numerical computation of convolutions (and their derivatives) of compactly supported kernels made the computing time explode, even with FFT. This is unfortunate, because smooth, compactly supported kernels are used for construction of confidence bands (cf. Assumption 2(a), even if, strictly speaking, the theory only allows the bandwidths to depend on $N(n)$ and, hence, does not allow for data-driven bandwidth selection). Moreover, as will be demonstrated below, the GL-estimators consistently tend to oversmooth, at least with sample sizes used in our simulations. Therefore, as a practical alternative, we propose to select the bandwidth for the estimator $\hat{f}_n$, given by formula (3), in another data-dependent way with the following version of the empirical risk minimization (ERM) principle, which is computationally much cheaper than the GL-method. Define

$$R(h) = ||\hat{f}_n - f||_2^2 - ||f||_2^2 = ||\hat{f}_n||_2^2 - 2\int_0^1 \hat{f}_n(x) f(x)\, dx.$$

Substituting $f(x) = -2xg'(x^2)$, we obtain, after a change of variable,

$$R(h) = ||\hat{f}_n||_2^2 + 2\int_0^1 \hat{f}_n(\sqrt{x})\, dg(x),$$

which can be estimated with

$$\hat{R}(h) = ||\hat{f}_n||_2^2 + 2\sum_{i=0}^{M-1} \hat{f}_n(\sqrt{x_i})\left[\tilde{g}(x_{i+1}) - \tilde{g}(x_i)\right],$$

where M is a positive integer, $0 = x_0, \ldots, x_M = 1$ are equally spaced points in $[0,1]$, and $\tilde{g}$ is some, independent of h, estimator of the intensity function g. The bandwidth h is chosen as the minimizer of $\hat{R}(h)$.

In the numerical experiment, we used $M = 2^{10}$ and the following kernel estimator of the function g:

$$\tilde{g}(x) = \frac{1}{nH} \sum_{i=1}^{N(n)} \left[K_{ep}\left(\frac{x - X_i^2}{H}\right) + K_{ep}\left(\frac{x + X_i^2}{H}\right) \right], \tag{5}$$

where K_{ep} is the Epanechnikov kernel and $H = 5(N(n)/n)^2 n^{-1/5}$, which proved to work well in a wide range of examples studied in simulations. Notice that $N(n)/n$ converges almost surely to $\int_0^1 g(x)\,dx$ and the function g is decreasing with $g(0) = \int_0^1 f(x)\,dx$ and $g(1) = 0$. This means that, for fixed $\int_0^1 f(x)\,dx$, the smaller $\int_0^1 g(x)\,dx$, the more rapid decrease of the function g. In that case, one should take a smaller bandwidth for better estimation of the function g. The minimization of $\hat{R}(h)$ was performed for $h = Cn^{-1/7}$ through a grid search with $C \in [0.2, 2.0]$ and with step 0.05 for C.

Data samples were generated from several intensity functions, but we only present selected results for the following four functions (taken from [16]) that represent typical behaviour of our procedures in various setups:

1. Beta(4,2): $f(x) = 20x^3(1-x)\mathbb{1}_{[0,1]}(x)$
2. Swapped Minerbo-Levy A (SML-A):
   $f(x) = 4x^2\mathbb{1}_{[0,0.5]}(x) + [2 - 4(1-x)^2]\mathbb{1}_{[0.5,1]}(x)$
3. Swapped Minerbo-Levy B (SML-B):
   $f(x) = 1.241(2x - x^2)^{-3/2}\exp\{1.21[1 - (2x - x^2)^{-1}]\}\mathbb{1}_{[0,1]}(x)$
4. Normal mixture (NM): $0.7 \cdot N(0.7, 0.08^2) + 0.3 \cdot N(0.35, 0.08^2)$.

Note that only Beta(4,2) belongs to $\mathcal{F}_{1,C}$, which guarantees the MISE convergence rate $n^{-2/5}$, according to Theorem 1. The other functions, being discontinuous at 1, fail to belong to $\mathcal{F}_{m,C}$. Nevertheless, all four functions satisfy the less restrictive conditions of Corollary 1, so that asymptotically valid confidence bands may be constructed for all of them. The mean numbers of the observed data points were 47.6%, 54.2%, 41.3%, and 38.6% of n, respectively for the functions from Beta(4,2) to NM.

The SML-B and NM functions are much more difficult to estimate in stereological problems than Beta(4,2) and SML-A. SML-B is hard mainly because of rapid local changes of the derivative, being close to zero over the rest of the support (step-like behaviour). For the NM function, in addition to rapid local changes of the derivative, problems with the coverage can be expected near to one, where the function is close to zero and steepy, because the confidence bands will be narrow there, due to the presence of $\tilde{q}_{N(n)}$ in the nominator of $l_n(t,x)$ (cf. Corollary 1). Additionally, for NM data is scarce, indeed, in that region because, due to the very nature of the SLW problem, the observed data are shifted to the left w.r.t. the original radii, and there is not much to be shifted from the neighbourhood of one, although Assumption 2(d) is formally satisfied.

For the GL-method, one has to select some value of the parameter η. Numerical experiments have shown that the bandwidths selected with $\eta = 0.5$ were

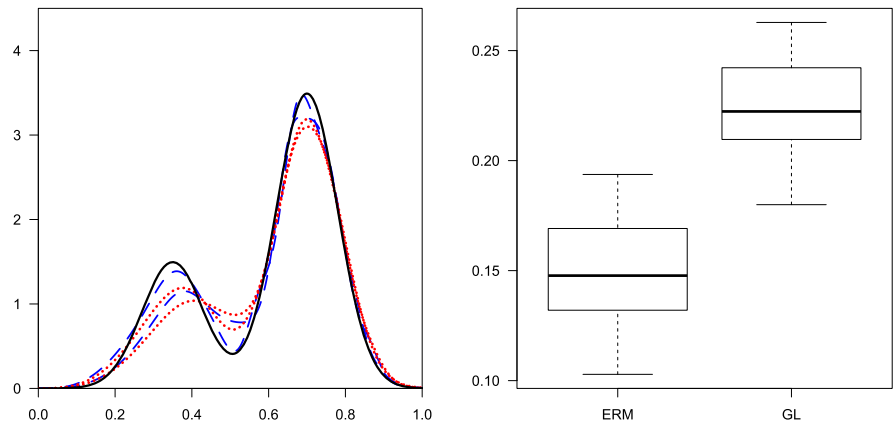

FIG 1. *Left: worst and best estimates (out of 10 data samples) of the NM intensity function (solid), obtained with GL (dotted) and ERM (dashed) method of bandwidth selection, with experiment size* $n = 10^5$. *Right:* L^2*-errors of ERM and GL based estimates for 10 data samples used in the left panel (medians, quartiles, extremes).*

always larger than those selected with $\eta = 10^{-2}$, and the increasing tendency continued to $\eta = 1.5$ and $\eta = 2$, for all generated data samples. On the other hand, the difference between using $\eta = 10^{-2}$ and $\eta = 10^{-4}$, say, was negligible. To compare the GL-method (with $\eta = 10^{-2}$) with the ERM method, we have run both (with the Gaussian kernel) on the same data and compared the estimators and their L^2 squared errors. Typically, GL-estimators tend to oversmooth and to have larger errors than those of the ERM-based estimators. Figure 1 illustrates that for the NM function, with $n = 10^5$ and estimators computed from 10 data samples. In view of the above discussion, only ERM method and estimators with the biweight kernel $K(x) = (15/16)(1 - x^2)^2 \mathbb{1}_{[-1,1]}(x)$ of order two in the estimator $\hat{f}_n$ will be further discussed below. Figures 2 and 3 illustrate the behaviour of the estimator $\hat{f}_n$ for Beta(4,2) and SML-B functions. For each function and each experiment size, 10 artificial data samples were generated and the estimator $\hat{f}_n$ with bandwidth selected with ERM principle was computed. Additionally, the best possible bandwidths were found, i.e. those that minimize the numerically computed L^2 distance between $\hat{f}_n$ and the true f. The best and worst cases (out of 10) are presented in the left panels of Figures 2 and 3. The right panels of those figures illustrate the performance of the ERM principle. For similar results on spectral-type and wavelet estimators, see [16, 12]. As seen in Figures 2 and 3, minimization of empirical risk produces good estimates of the unfolded intensity function, close to those with the optimal bandwidth.

The last part of this section presents the results of our simulation study of the asymptotic confidence bands given in Corollary 1. The asymptotic theory applies to confidence bands constructed on intervals $[a, b]$ with any $0 < a < b < 1$. In finite samples, however, a and b should not be chosen too close to the bound-

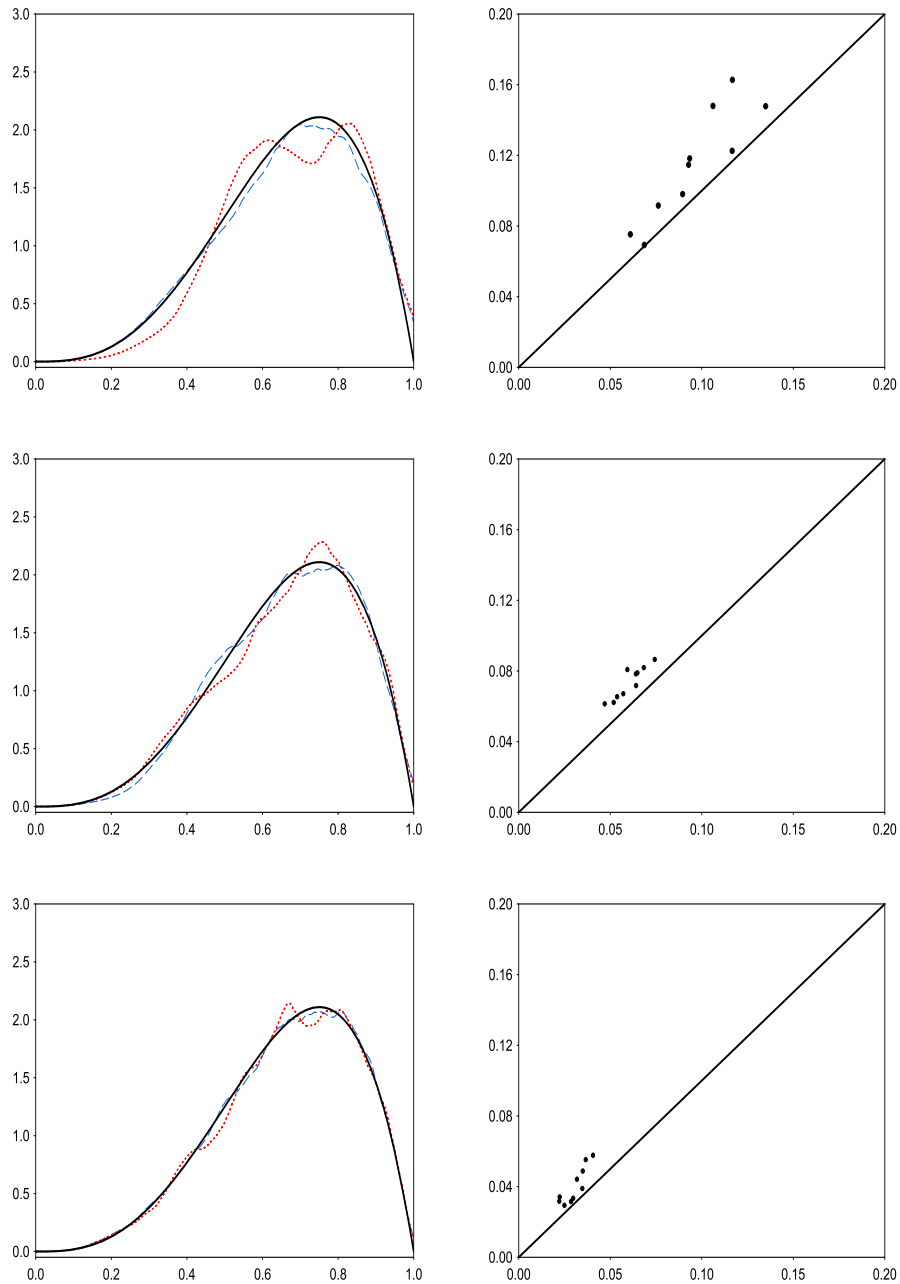

FIG 2. *Worst (dotted) and best (dashed) reconstruction (out of 10 data samples) of Beta(4,2) intensity function (solid line) for the experiment size, from top:* $n = 10^4$, $n = 10^5$, $n = 10^6$. *On the right side the scatterplots of* L^2 *error of ERM solutions versus those of the best possible solutions.*

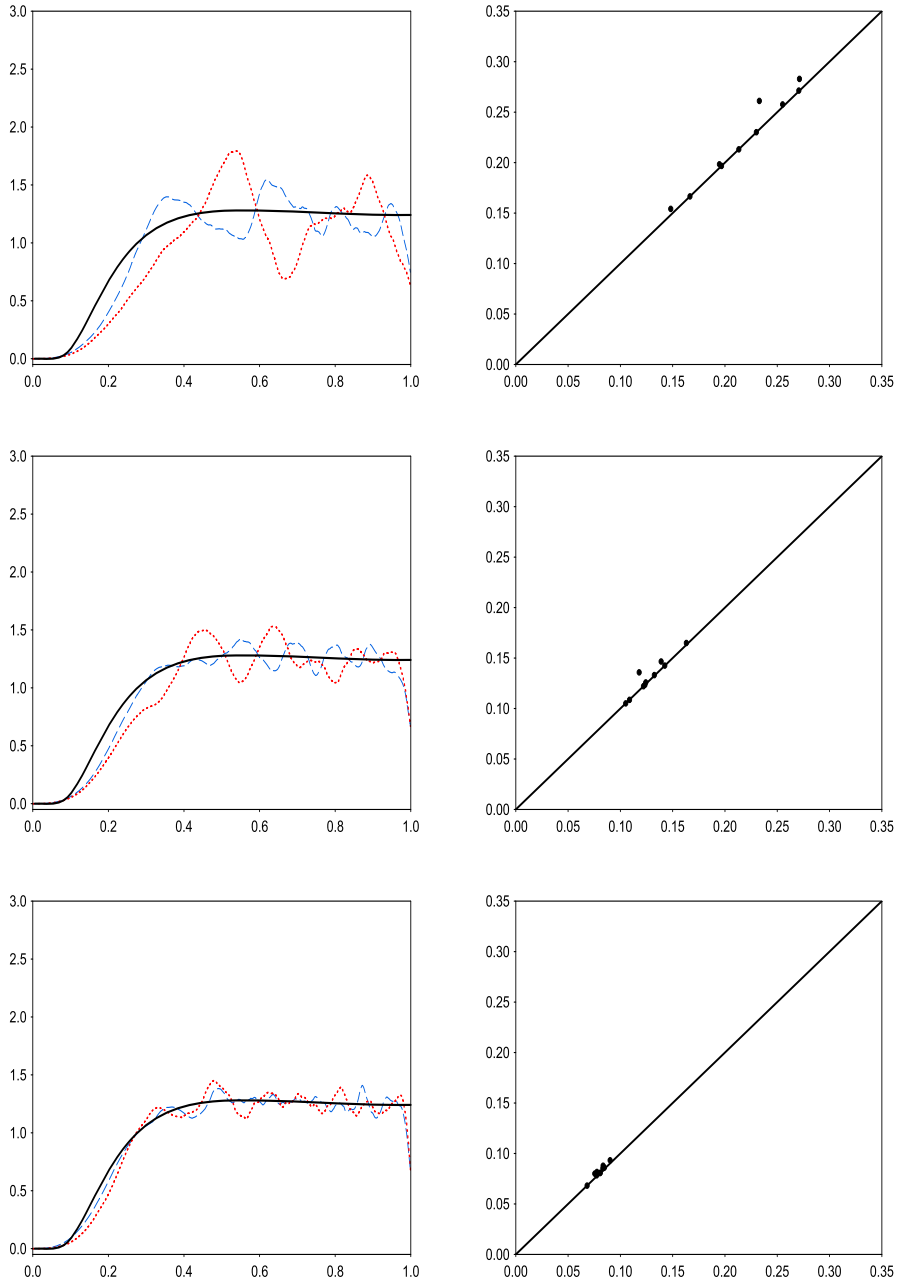

FIG 3. *Similar to Figure 2, but for the SML-B intensity function.*

aries. Problems with actual coverage probability can be expected because of at least two reasons: wrong width of the band and/or bad behaviour of the central estimator. The half-width $l_n(t,x)$ of the band is proportional to the estimate $\tilde{q}_{N(n)}$ of the density $q(\cdot) = g(\cdot)/c$ (cf. Corollary 1). With $f \in \mathcal{F}_{m,C}$, symmetrization allows for reliable estimation of $q(\cdot)$ near to zero, so that the bands width may be expected reasonable there (provided h is chosen correctly). In the vicinity of one, however, problems with estimation of $q(\cdot)$ should be expected, due to well-known boundary effects. Moreover, since $q(1) = 0$, the bands can become too narrow in the vicinity of one. On the other hand, the central estimator $\hat{f}_n$ may be expected to cause problems at both ends of the interval $[0,1]$ due to boundary effects, because it is based on the estimator of the derivative of g and symmetrization does not necessarily help in this case. Problems may be more serious near to zero, due to the additional square transformation of the argument (cf. formula (3)). In effect, to fully eliminate the boundary effects, one would have to concider only a and b such that $\sqrt{h} < a < b < \sqrt{1-h}$, which may be quite restrictive in finite samples. To which extent these restrictions can be relaxed, may depend on the estimated function.

In the first stage of our numerical experiment, all confidence bands were constructed on the interval $[0.1, 0.9]$ and with the bandwidth $h = 0.85h^*$, where h^* was chosen according to the ERM procedure. For rapidly changing functions SML-B and NM, however, the coverage probabilities were too low, and we were forced to use $h = 0.75h^*$ in those cases. Alternatively, one can use $h = 0.85h^*$ on a shorter interval. As $\tilde{q}_{N(n)}$, we used the kernel estimator (5), with $N(n)$ instead of n in the normalizing factor.

Table 1 shows the simulated coverage probabilities and the confidence bands mean widths for all considered intensity functions f, for three values of the experiment size n: 10 000, 30 000, 100 000, and for three levels of nominal coverage probability. The results are based on 1 000 simulation runs. Note that, out of all scenarios for SML-B, only the last row of the table (approximately) meets the restriction $\sqrt{h} < a$ and, at the same time, produces coverage probabilities very close to nominal ones. The remaining functions are much less problematic in this respect.

Figures 4 and 5 show some typical examples of 80% and 95% confidence bands for $n = 10^4$ and $n = 10^5$. It is seen in those figures and in Table 1 that the bandwidth selected with the ERM procedure may also be used, after multiplication by a shrinkage factor, in the construction of confidence bands. The difficulty of the SML-B case is reflected in larger sample sizes needed to obtain reasonable estimates and in too low actual coverage probabilities in smaller samples. Similar effect, although less pronounced, can be seen in case of the NM function.

6. Discussion and conclusions

The proposed kernel-type estimator with an automatic bandwidth selection and the related confidence bands provide the first solution to the problem of uniform

TABLE 1

Simulated coverage percentages (Cov.) and mean widths of the confidence bands (Mean width = mean area/length) for various intensity functions f, various experiment sizes n and various intervals (S) and shrinkage factors applied to h^*. Mean h is averaged over 1 000 simulations, run to empirically approximate each of the coverage percentages. Approximate standard errors for simulated coverage percentages are 1.3%, 0.9%, *and* 0.7%, *respectively for nominal coverage* 80%, 90%, *and* 95%.

			Nominal coverage					
			80%		90%		95%	
Intensity	n	Mean h	Cov.	Mean width	Cov.	Mean width	Cov.	Mean width
$h = 0.85h^*$, $S = [0.1, 0.9]$								
Beta(4.2)	10 000	0.180	83	0.71	93	0.82	98	0.92
	30 000	0.145	84	0.59	94	0.67	97	0.75
	100 000	0.114	84	0.48	92	0.54	98	0.60
SML-A	10 000	0.234	80	0.50	93	0.59	97	0.67
	30 000	0.189	76	0.42	89	0.48	96	0.54
	100 000	0.148	76	0.34	88	0.39	95	0.43
$h = 0.75h^*$, $S = [0.1, 0.9]$								
NM	10 000	0.105	73	1.40	87	1.56	95	1.70
	30 000	0.085	75	1.15	89	1.28	95	1.40
	100 000	0.066	81	0.95	90	1.06	95	1.14
SML-B	10 000	0.120	64	1.29	78	1.45	87	1.60
	30 000	0.096	68	1.08	81	1.21	88	1.33
	100 000	0.076	66	0.87	79	0.96	87	1.06
$h = 0.85h^*$, $S = [0.25, 0.9]$								
SML-B	10 000	0.136	63	1.18	74	1.34	84	1.50
	30 000	0.109	70	0.98	81	1.11	90	1.23
	100 000	0.086	79	0.79	89	0.88	96	0.97

interval estimation in the Spektor-Lord-Willis problem of stereology, formulated as a Poisson inverse problem. As seen in Proposition 1, the smoothness class $\mathcal{F}_{m,C}$ is essentially the Sobolev class $W(m, L)$, with some local restrictions in an arbitrarily small vicinity of zero. Balls with radii below detectability limit in a given experiment are, however, not observable anyway, so that such local restrictions seem acceptable from the applied point of view. As shown in Section 2, estimating f in the standard L^2-norm is equivalent to estimating g' in the weighted L^2-norm, with the weight $\sqrt{x}$. This weight does not influence the minimax rate, which is $n^{-2m/(2m+3)}$, as one would expect with the standard L^2-norm. Recall, however, that g has $m-1$ zero derivatives at zero, which makes estimation of g' in the vicinity of zero an easy task. Hence, the estimation error in $(\epsilon, 1]$ dominates the global error and the weight does not influence the minimax rates.

Relatively large samples needed for the applicability of asymptotic confidence bands motivate studying alternative bootstrap constructions. This is also

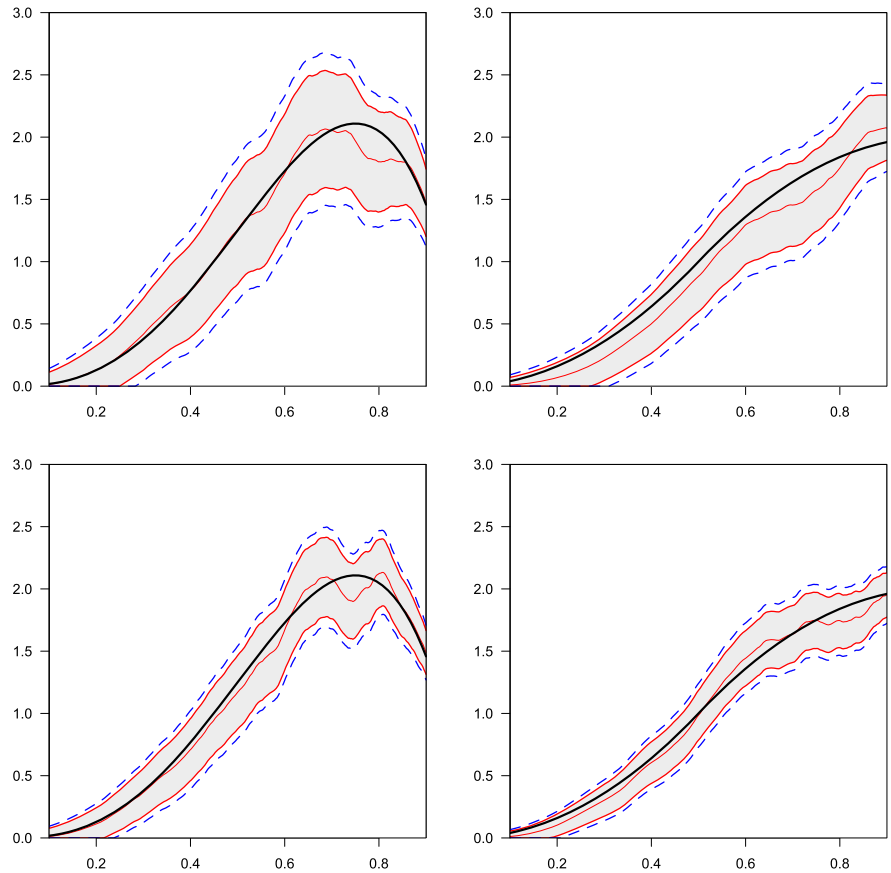

FIG 4. *Estimate* $\hat{f}_n$ *along with associated* 80% *(shaded) and* 95% *(– – –) nominal coverage probability confidence bands of the intensity functions Beta(4,2) (left) and SML-A (right) for the experiment size* $n = 10^4$ *(top) and* $n = 10^5$ *(bottom). The thick solid lines represent the true functions.*

attractive from theoretical point of view because, strictly speaking, asymptotic confidence bands based on the standard Bickel-Rosenblatt theorem do not allow for data-driven choice of the bandwidth. This restriction can be relaxed only at the cost of serious additional theoretical effort, as in [18], or in [9], where an approach based on, so-called, anti-concentration property of the supremum of a relevant Gaussian process was proposed, along with a Gaussian multiplier bootstrap version of the Lepski's method and a corresponding construction of confidence bands for densities. Adaptation of those techniques to our setup and empirical verification of the performance of such (again, computationally very costly) confidence bands is, however, beyond the scope of this article and may be the subject of a separate project.

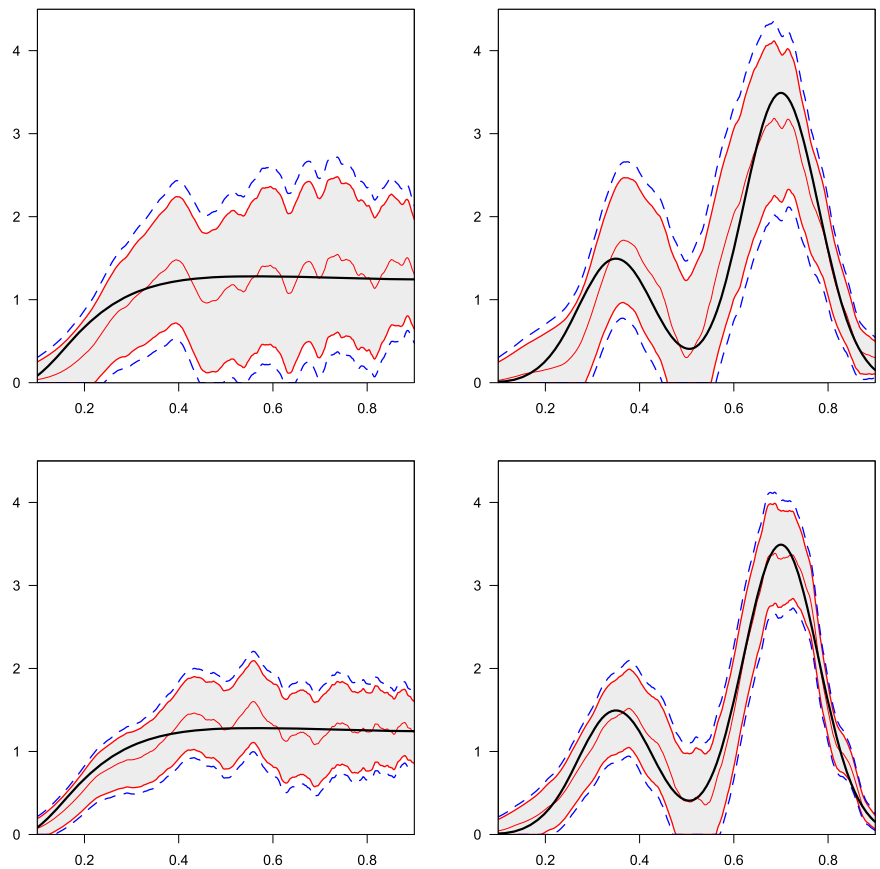

FIG 5. *Similar to Figure 4, but for the intensity functions SML-B (left) and NM (right).*

Appendix A: Proof of Proposition 2

We closely follow the ideas from the proof of Proposition 3 in [15] and Theorem 2 in [28]. For any $h \in \mathcal{H}$,

$$\|\hat{p'}_{GL} - p'\|_* \leq \|\hat{p'}_{GL} - \hat{p'}_{\hat{h},h}\|_* + \|\hat{p'}_{\hat{h},h} - \hat{p'}_h\|_* + \|\hat{p'}_h - p'\|_* =: A_1 + A_2 + A_3,$$

with

$$A_1 \leq A(h) + \frac{\chi\sqrt{N(n)}}{n\hat{h}^{3/2}} \quad \text{and} \quad A_2 \leq A(\hat{h}) + \frac{\chi\sqrt{N(n)}}{nh^{3/2}}.$$

By definition of $\hat{h}$,

$$A_1 + A_2 \leq 2A(h) + \frac{2\chi\sqrt{N(n)}}{nh^{3/2}}.$$

Set

$$\xi_n(h) = \sup_{h' \in \mathcal{H}} \left\{ \|(\hat{p'}_{h,h'} - \mathbb{E}\hat{p'}_{h,h'}) - (\hat{p'}_{h'} - \mathbb{E}\hat{p'}_{h'})\|_2 - \frac{\chi\sqrt{N(n)}}{n(h')^{3/2}} \right\}_+ .$$

Then, because $\|\hat{p'}_{h,h'} - \hat{p'}_{h'}\|_* \le \|\hat{p'}_{h,h'} - \hat{p'}_{h'}\|_2$,

$$A_1 + A_2 \le 2\xi_n(h) + 2 \sup_{h' \in \mathcal{H}} \|\mathbb{E}\hat{p'}_{h,h'} - \mathbb{E}\hat{p'}_{h'}\|_2 + \frac{2\chi\sqrt{N(n)}}{nh^{3/2}}.$$

For the middle term, recall that $\mathbb{E}\hat{p'}_{h'} = K_{h'} \star p'$ and evaluate

$$\begin{aligned}
\mathbb{E}\hat{p'}_{h,h'} &= \mathbb{E}\left[\mathbb{E}\left(\hat{p'}_{h,h'} | N(n)\right)\right] \\
&= \mathbb{E}\left[\frac{N(n)}{n\|g\|_1}\left(\int (K_h \star K_{h'})'(x-u)g(u)\,du + \int (K_h \star K_{h'})'(x+u)g(u)\,du\right)\right] \\
&= \int (K_h \star K_{h'})'(x-u)g(u)\,du + \int (K_h \star K_{h'})'(x-u)g(-u)\,du \\
&= \int (K_h \star K_{h'})'(x-u)p(u)\,du = K_h \star K_{h'} \star p'.
\end{aligned}$$

Then,

$$\begin{aligned}
A_1 + A_2 &\le 2\xi_n(h) + 2 \sup_{h' \in \mathcal{H}} \|K_h \star K_{h'} \star p' - K_{h'} \star p'\|_2 + \frac{2\chi\sqrt{N(n)}}{nh^{3/2}} \\
&\le 2\xi_n(h) + 2\|K\|_1 \|K_h \star p' - p'\|_2 + \frac{2\chi\sqrt{N(n)}}{nh^{3/2}}.
\end{aligned}$$

As a byproduct, it is seen that $A(h)$ represents, up to the factor $\|K\|_1$, the approximation term. From $(a+b+c)^2 \le 3a^2 + 3b^2 + 3c^2$, one gets

$$\mathbb{E}(A_1 + A_2)^2 \le 12\mathbb{E}\xi_n^2(h) + 12\|K\|_1^2 \|K_h \star p' - p'\|_2^2 + \frac{12\chi^2\|g\|_1}{nh^3}.$$

For A_3, we obtained in the proof of Theorem 1 in Section 2

$$\mathbb{E}A_3^2 \le \frac{4\|g\|_1\|K'\|_2^2}{nh^3} + \|K_h \star p' - p'\|_2^2.$$

Finally, using $(a+b)^2 \le 2a^2 + 2b^2$,

$$\begin{aligned}
\mathbb{E}\|\hat{p'}_{GL} - p'\|_*^2 &\le 2\mathbb{E}(A_1 + A_2)^2 + 2\mathbb{E}A_3^2 \\
&\le 2\left[4 + 12(1+\eta)^2(1 + \|K\|_1)^2\right]\left\{\|K_h \star p' - p'\|_2^2 + \frac{\|g\|_1\|K'\|_2^2}{nh^3}\right\} \\
&\quad + 24\mathbb{E}\xi_n^2(h).
\end{aligned}$$

This will give the thesis because of the following lemma.

LEMMA **1.** *If $\mathcal{H} \subset \{D^{-1} : D = 1, \dots, \lfloor (\delta n)^{1/2} \rfloor\}$ with some $\delta > 0$, then there exists a constant $C = C(\delta, \eta, \|K\|_2, \|K'\|_2, \|g\|_1, \|g\|_\infty)$ such that, for each $h \in \mathcal{H}$ and for all n, $\mathbb{E}\xi_n^2(h) \le Cn^{-1}$.*

Proof. For any $h \in \mathcal{H}$,

$$\xi_n(h) \le \sup_{h' \in \mathcal{H}} \left\{ \|\hat{p'}_{h,h'} - \mathbb{E}\hat{p'}_{h,h'}\|_2 + \|\hat{p'}_{h'} - \mathbb{E}\hat{p'}_{h'}\|_2 - \frac{\chi\sqrt{N(n)}}{n(h')^{3/2}} \right\}_+ ,$$

$$\|\hat{p'}_{h,h'} - \mathbb{E}\hat{p'}_{h,h'}\|_2 = \|K_h \star \hat{p'}_{h'} - K_h \star K_{h'} * p'\|_2 \le \|K\|_1 \|\hat{p'}_{h'} - \mathbb{E}\hat{p'}_{h'}\|_2,$$

and, consequently,

$$\begin{aligned}
\xi_n(h) &\le \sup_{h' \in \mathcal{H}} \left\{ (1 + \|K\|_1)\|\hat{p'}_{h'} - \mathbb{E}\hat{p'}_{h'}\|_2 - \frac{\chi\sqrt{N(n)}}{n(h')^{3/2}} \right\}_+ \\
&= (1 + \|K\|_1) \sup_{h' \in \mathcal{H}} \left\{ \|\hat{p'}_{h'} - \mathbb{E}\hat{p'}_{h'}\|_2 - \frac{(1+\eta)\|K'\|_2\sqrt{N(n)}}{n(h')^{3/2}} \right\}_+ \\
&=: (1 + \|K\|_1) S_n.
\end{aligned}$$

With any $\alpha \in (0,1)$, we then have $\mathbb{E}\xi_n^2(h) \le (1 + \|K\|_1)^2(A + B)$, with

$$\begin{aligned}
A &:= \mathbb{E}\left[S_n^2 \mathbb{1}(N(n) \le (1-\alpha)^2 n\|g\|_1)\right], \\
B &:= \mathbb{E}\left[S_n^2 \mathbb{1}(N(n) > (1-\alpha)^2 n\|g\|_1)\right].
\end{aligned}$$

To handle A, we use again $(a_1 + \cdots + a_\ell)^2 \le \ell(a_1^2 + \cdots + a_\ell^2)$ and obtain

$$\begin{aligned}
S_n^2 &\le \sup_{h \in \mathcal{H}} \|\hat{p'}_h - \mathbb{E}\hat{p'}_h\|_2^2 \le 2 \sup_{h \in \mathcal{H}} \|\hat{p'}_h\|_2^2 + 2 \sup_{h \in \mathcal{H}} \|\mathbb{E}\hat{p'}_h\|_2^2 \\
&\le 2 \sup_{h \in \mathcal{H}} \int \left[\frac{1}{n} \sum_{i=1}^{N(n)} \left[K_h'(x - X_i^2) + K_h'(x + X_i^2)\right] \right]^2 dx + 2 \sup_{h \in \mathcal{H}} \|K_h \star p'\|_2^2 \\
&\le 2 \sup_{h \in \mathcal{H}} \frac{2N(n)}{n^2} \int \sum_{i=1}^{N(n)} \left(K_h'(x - X_i^2)^2 + K_h'(x + X_i^2)^2\right) dx + 2 \sup_{h \in \mathcal{H}} \|K_h \star p'\|_2^2 \\
&\le 8 \sup_{h \in \mathcal{H}} \frac{N^2(n)\|K'\|_2^2}{n^2 h^3} + 8 \sup_{h \in \mathcal{H}} \frac{\|g\|_1^2 \|K'\|_2^2}{h^3} \\
&\le 8\|K'\|_2^2 (\delta n)^{3/2} \left(\frac{N^2(n)}{n^2} + \|g\|_1^2 \right),
\end{aligned}$$

because $h^3 \ge (\delta n)^{-3/2}$. Further, since $N(n)/n \le \|g\|_1$, one obtains with $C_0 = 16\|K'\|_2^2 (\delta n)^{3/2} \|g\|_1^2$

$$A \le C_0 P\left(N(n) \le (1-\alpha)^2 n\|g\|_1\right) \le C_0 \exp[-\alpha'\|g\|_1 n],$$

with some $\alpha'(\alpha)$, as in [28, Additional File 2]. This finally gives, for all n,

$$A \le C_A n^{-1},$$

with a constant $C_A = C_A(\|g\|_1, \|K'\|_2, \delta, \alpha)$.

Now we deal with B, fixing the previous value of α to $\alpha = \min(\eta/2, 1/4)$, which implies $(1+\eta)(1-\alpha) \geq 1+\eta/4$ and

$$\begin{aligned}
B &\leq \mathbb{E}\left[\sup_{h\in\mathcal{H}}\left\{\|\hat{p'}_h - \mathbb{E}\hat{p'}_h\|_2 - \frac{(1+\eta/4)\|K'\|_2\|g\|_1^{1/2}}{n^{1/2}h^{3/2}}\right\}_+^2\right] \\
&= \int_0^\infty P\left(\sup_{h\in\mathcal{H}}\left\{\|\hat{p'}_h - \mathbb{E}\hat{p'}_h\|_2 - \frac{(1+\eta/4)\|K'\|_2\|g\|_1^{1/2}}{n^{1/2}h^{3/2}}\right\}_+^2 \geq x\right) dx \\
&\leq \sum_{h\in\mathcal{H}}\int_0^\infty P\left(\left\{\|\hat{p'}_h - \mathbb{E}\hat{p'}_h\|_2 - \frac{(1+\eta/4)\|K'\|_2\|g\|_1^{1/2}}{n^{1/2}h^{3/2}}\right\}_+^2 \geq x\right) dx.
\end{aligned}$$

Set

$$U(x) = \hat{p'}_h(x) - \mathbb{E}\hat{p'}_h(x) = \frac{1}{n}\int [K_h'(x-t) + K_h'(x+t)]\, dN_{ng}(t) - (K_h \star p')(x),$$

where $N_{ng} = \sum_{i=1}^{N(n)} \delta_{X_i^2}$ is the observed Poisson process. In any Hilbert space, $\|u\| = \sup_{\|a\|\leq 1} |\langle a, u\rangle|$. Hence, if $\mathcal{A}$ is a countable, dense subset of the unit ball in $L^2(\mathbb{R})$, we have, because of the symmetry of K,

$$\begin{aligned}
\|U\|_2 &= \sup_{a\in\mathcal{A}}\left|\int a(x)U(x)\, dx\right| \\
&= \sup_{a\in\mathcal{A}}\left|\int a(x)\left[\frac{1}{n}\int (K_h'(x-t) + K_h'(x+t))\, dN_{ng}(t)\right.\right. \\
&\qquad\qquad \left.\left. - \int K_h'(x-t)p(t)\, dt\right] dx\right| \\
&= \sup_{a\in\mathcal{A}}\left|\int a(x)\left[\frac{1}{n}\int K_h'(x-t)\, dN_{ng}(t) - \int K_h'(x-t)g(t)\, dt\right.\right. \\
&\qquad\qquad \left.\left. + \frac{1}{n}\int K_h'(x+t)\, dN_{ng}(t) - \int K_h'(x-t)g(-t)\, dt\right] dx\right| \\
&\leq 2\sup_{a\in\mathcal{A}}\left|\int \frac{a(x)}{n}\left[\int K_h'(x-t)(dN_{ng}(t) - ng(t)dt)\right] dx\right| \\
&= \sup_{a\in\mathcal{A}}\left|\int \psi_a(t)(dN_{ng}(t) - ng(t)dt)\right|,
\end{aligned}$$

with

$$\psi_a(t) = \int \frac{2a(x)}{n} K_h'(x-t)\, dx.$$

For any t,

$$\psi_a^2(t) \leq \frac{4}{n^2}\int a^2(x)\, dx \int [K_h'(x-t)]^2\, dx = \frac{4\|K'\|_2^2}{n^2h^3}.$$

Set $b = 2\|K'\|_2/(nh^{3/2})$. Then $\|\psi_a\|_\infty \le b$ and

$$\mathbb{E}\|U\|_2^2 = \int \mathbb{E}U^2(x)\,dx = \int \operatorname{var}\left[\hat{p'}_h(x)\right] dx \le \frac{4\|g\|_1\|K'\|_2^2}{nh^3},$$

as in the proof of Theorem 1 in Section 2, which gives

$$\mathbb{E}\|U\|_2 \le \frac{2\|g\|_1^{1/2}\|K'\|_2}{n^{1/2}h^{3/2}}.$$

Also,

$$\begin{aligned} v &:= \sup_{a\in\mathcal{A}} \int \psi_a^2(t) ng(t)\,dt = \sup_{a\in\mathcal{A}} \int \left[\int \frac{2a(x)}{n} K_h'(x-t)\,dx\right]^2 ng(t)\,dt \\ &\le \frac{4}{n} \sup_{a\in\mathcal{A}} \int \int a^2(x)|K_h'(x-t)|\,dx \|K_h'\|_1 g(t)\,dt \\ &= \frac{4\|K_h'\|_1}{n} \int a^2(x) \int g(t)|K_h'(x-t)|\,dt\,dx \le \frac{4\|K_h'\|_1^2\|g\|_\infty}{n} = \frac{4\|K'\|_1^2\|g\|_\infty}{nh^2}. \end{aligned}$$

Since

$$\|\hat{p'}_h - \mathbb{E}\hat{p'}_h\|_2 \le \sup_{a\in\mathcal{A}} \left| \int \psi_a(t)(dN_{ng}(t) - ng(t)dt) \right|,$$

Corollary 2 in [27] gives, for any $\epsilon > 0$ and $u > 0$,

$$\begin{aligned} P\Bigg[\|\hat{p'}_h - \mathbb{E}\hat{p'}_h\|_2 \ge (1+\epsilon)\frac{2\|g\|_1^{1/2}\|K'\|_2}{n^{1/2}h^{3/2}} + \sqrt{\frac{48\|K'\|_1^2\|g\|_\infty u}{nh^2}} \\ + \left(\frac{5}{4} + \frac{32}{\epsilon}\right)\frac{2\|K'\|_2}{nh^{3/2}}u\Bigg] \le \mathrm{e}^{-u}. \end{aligned}$$

Setting $u = x + L_h$ with $x > 0$ gives

$$\begin{aligned} P\Bigg[\|\hat{p'}_h - \mathbb{E}\hat{p'}_h\|_2 - M_h \ge \sqrt{\frac{48\|K'\|_1^2\|g\|_\infty x}{nh^2}} \\ + \left(\frac{5}{4} + \frac{32}{\epsilon}\right)\frac{2\|K'\|_2}{nh^{3/2}}x\Bigg] \le \mathrm{e}^{-x}\mathrm{e}^{-L_h}, \end{aligned}$$

with

$$M_h = (1+\epsilon)\frac{2\|g\|_1^{1/2}\|K'\|_2}{n^{1/2}h^{3/2}} + \sqrt{\frac{48\|K'\|_1^2\|g\|_\infty L_h}{nh^2}} + \left(\frac{5}{4} + \frac{32}{\epsilon}\right)\frac{2\|K'\|_2}{nh^{3/2}}L_h.$$

One has

$$\begin{aligned} &\mathbb{E}\left(\sup_{h\in\mathcal{H}}\left[\|\hat{p'}_h - \mathbb{E}\hat{p'}_h\|_2 - M_h\right]_+^2\right) \\ &\qquad \le \sum_{h\in\mathcal{H}} \int_0^\infty P\left(\left[\|\hat{p'}_h - \mathbb{E}\hat{p'}_h\|_2 - M_h\right]_+^2 > v\right) dv. \end{aligned} \tag{6}$$

In the last integral substitute

$$\sqrt{v} = \sqrt{48\|g\|_\infty}\|K'\|_1 \frac{\sqrt{u}}{n^{1/2}h^{3/2}} + 2\left(\frac{5}{4} + \frac{32}{\epsilon}\right)\|K'\|_2 \frac{u}{nh^{3/2}}.$$

The right hand side of (6) can then be bounded by

$$\sum_{h\in\mathcal{H}} \mathrm{e}^{-L_h}\left(\frac{D_1}{nh^3} + \frac{D_2}{n^{3/2}h^3} + \frac{D_3}{n^2h^3}\right),$$

with constants D_1, D_2, D_3 that depend only on $\|K'\|_1$, $\|K'\|_2$, $\|g\|_\infty$, and ϵ. Set $L_h = \theta^2/h$. Then, since $h^{-1} \le (\delta n)^{1/2}$,

$$\begin{aligned} M_h \le &(1+\epsilon+\theta\sqrt{12})\frac{2\max\{\|g\|_1^{1/2}, \|g\|_\infty^{1/2}\}\max\{\|K'\|_1, \|K'\|_2\}}{n^{1/2}h^{3/2}} \\ &+ 2\left(\frac{5}{4} + \frac{32}{\epsilon}\right)\frac{\|K'\|_2\theta^2\delta^{1/2}}{n^{1/2}h^{3/2}}. \end{aligned}$$

After suitable choice of ϵ and θ, one gets

$$M_h \le \frac{(1+\eta/4)\|K'\|_2\|g\|_1^{1/2}}{n^{1/2}h^{3/2}}$$

and, hence, with $D = \max\{D_1, D_2, D_3\}$,

$$B \le \frac{D}{n}\sum_{i=1}^{\infty} \mathrm{e}^{-\theta^2 i} i^3 \le C_B n^{-1},$$

with some constant $C_B = C_B(\|K'\|_1, \|K'\|_2, \|g\|_\infty, \|g\|_2)$. This completes the proof of Lemma 1 and of Proposition 2.

Recall from Section 2 that g is nonincreasing and $g(0) = \int f(x)\,dx$. This implies that $\|g\|_\infty = \|f\|_1$. Also, $\|g\|_1 \le \|f\|_1$, so that all the constants will be uniform over the function class, if we impose a limit $\|f\|_1 \le D$, say, in addition to the definition of $\mathcal{F}_{n,C}$. Note that if a function f satisfies the conditions of Proposition 1, then automatically $\|f\|_1$ has an upper limit.

Acknowledgments

This work was partially supported by the Faculty of Applied Mathematics AGH UST statutory tasks and dean grant for PhD students and young researchers within subsidy of Ministry of Science and Higher Education. The computational part of this research was supported in part by PL-Grid Infrastructure. The authors thank the reviewers for helpful suggestions of improvements, especially those related to adaptivity of the estimator and to the simulation experiment.

References

[1] Antoniadis, A. and Bigot, J. (2006). Poisson inverse problems. *The Annals of Statistics* **34** 2132–2158. MR2291495

[2] Barthel, M., Klimanek, P. and Stoyan, D. (1985). Stereological substructure analysis in hot-deformed metals from TEM-images. *Czech. Journal of Physics B* **35** 265–268.

[3] Bickel, P. J. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *The Annals of Statistics* **1** 1071–1095. MR0348906

[4] Birke, M., Bissantz, N. and Holzmann, H. (2010). Confidence bands for inverse regression models. *Inverse Problems* **26** article 115020. MR2732912

[5] Bissantz, N. and Birke, M. (2009). Asymptotic normality and confidence intervals for inverse regression models with convolution-type operators. *Journal of Multivariate Analysis* **100** 2364–2375. MR2560377

[6] Bissantz, N., Dümbgen, L., Holzmann, H. and Munk, A. (2007). Nonparametric confidence bands in deconvolution density estimation. *Journal of the Royal Statistical Society, Ser. B* **69** 483–506. MR2323764

[7] Bissantz, N. and Holzmann, H. (2008). Statistical inference for inverse problems. *Inverse Problems* **24** article 034009. MR2421946

[8] Bissantz, N., Holzmann, H. and Proksch, K. (2014). Confidence regions for images observed under the Radon transform. *Journal of Multivariate Analysis* **128** 86–107. MR3199830

[9] Chernozhukov, V., Chetverikov, D. and Kato, K. (2014). Anticoncentration and honest, adaptive confidence bands. *The Annals of Statistics* **42** 1787–1818. MR3262468

[10] Chiu, S. N., Stoyan, D., Kendall, W. S. and Mecke, J. (2013). *Stochastic Geometry and its Applications*. Wiley, Chichester. MR3236788

[11] Ćmiel, B. (2010). An adaptive wavelet shrinkage approach to the Spektor-Lord-Willis problem. *Journal of Multivariate Analysis* **101** 1458–1470. MR2609506

[12] Ćmiel, B. (2012). Poisson intensity estimation for the Spektor-Lord-Willis problem using a wavelet shrinkage approach. *Journal of Multivariate Analysis* **112** 194–206. MR2957296

[13] Delaigle, A., Hall, P. and Jamshidi, F. (2015). Confidence bands in nonparametric errors-in-variables regression. *Journal of the Royal Statistical Society, Ser. B* **77** 149–169. MR3299403

[14] Dony, J. and Einmahl, U. (2006). Weighted uniform consistency of kernel density estimators with general bandwidth sequences. *Electronic Journal of Probability* **11** 844–859. MR2261055

[15] Doumic, M., Hoffmann, M., Reynaud-Bouret, P. and Rivoirard, V. (2012). Nonparametric estimation of the division rate of a size-structured population. *SIAM Journal on Numerical Analysis* **50** 925–950. MR2914292

[16] Dudek, A. and Szkutnik, Z. (2008). Minimax unfolding of the spheres

size distribution from linear sections. *Statistica Sinica* **18** 1063–1080. MR2440404

[17] Eubank, R. L. and Speckman, P. L. (1993). Confidence bands in nonparametric regression. *Journal of the American Statistical Association* **88** 1287–1301. MR1245362

[18] Giné, E. and Nickl, R. (2010). Confidence bands in density estimation. *The Annals of Statistics* **38** 1122–1170. MR2604707

[19] Giné, E. and Nickl, R. (2016). *Mathematical Foundations of Infinite-Dimensional Statistical Models. Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, New York. MR3588285

[20] Goldenshluger, A. and Lepski, O. (2011). Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *The Annals of Statistics* **39** 1608–1632. MR2850214

[21] Han, J. H. and Kim, D. Y. (1998). Determination of three-dimensional grain size distribution by linear intercept measurement. *Acta Materialia* **46** 2021–2028.

[22] Hilliard, J. E. and Lawson, L. R. (2003). *Stereology and Stochastic Geometry*. Kluwer, Dordrecht.

[23] Lord, G. W. and Willis, T. F. (1951). Calculation of air bubble distribution from results of a Rosiwal traverse of aerated concrete. *A.S.T.M. Bulletin* **56** 177–187.

[24] Lounici, K. and Nickl, R. (2011). Global uniform risk bounds for wavelet deconvolution estimator. *The Annals of Statistics* **39** 201–231. MR2797844

[25] Proksch, K., Bissantz, N. and Dette, H. (2015). Confidence bands for multivariate and time dependent inverse regression models. *Bernoulli* **21** 144–175. MR3322315

[26] Reiss, R. D. (1993). *A Course on Point Processes*. Springer, New York. MR1199815

[27] Reynaud-Bouret, P. (2003). Adaptive estimation of the intensity of inhomogeneous Poisson processes via concentration inequalities. *Probability Theory and Related Fields* **126** 103–153.

[28] Reynaud-Bouret, P., Rivoirard, V., Grammont, F. and Tuleau-Malot, C. (2014). Goodness-of-fit tests and nonparametric adaptive estimation for spike train analysis. *Journal of Mathematical Neuroscience* **4:3** 1–41. MR3197017

[29] Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London. MR848134

[30] Spektor, A. G. (1950). Analysis of distribution of spherical particles in non-transparent structures. *Zavodskaja Laboratorija* **16** 173–177.

[31] Szkutnik, Z. (2000). Unfolding intensity function of a Poisson process in models with approximately specified folding operator. *Metrika* **52** 1–26. MR1791921

[32] Szkutnik, Z. (2005). B-splines and discretization in an inverse problem for Poisson processes. *Journal of Multivariate Analysis* **93** 198–221. MR2119771

[33] Szkutnik, Z. (2007). Unfolding spheres size distribution from linear sec-

tions with B-splines and EMDS algorithm. *Opuscula Mathematica* **27** 151–165. MR2320844

[34] Szkutnik, Z. (2010). A note on minimax rates of convergence in the Spektor-Lord-Willis problem. *Opuscula Mathematica* **30** 203–207. MR2669132

[35] Tsybakov, A. B. (2009). *Introduction to Nonparametric Estimation.* Springer, New York. MR2724359

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