The Annals of Statistics
1995, Vol. 23, No. 5, 1518–1542

ISOTONIC ESTIMATION AND RATES OF CONVERGENCE IN WICKSELL'S PROBLEM

By Piet Groeneboom and Geurt Jongbloed

Delft University of Technology

It is shown that, in the nonparametric setting for the so-called Wicksell problem, the distribution function of the squared radii of the balls cannot be estimated at a rate faster than $n^{-1/2}\sqrt{\log n}$. We present an isotonic estimator of the distribution function which attains this rate and derive its asymptotic (normal) distribution. It is shown that the variance of this limiting distribution is exactly half the asymptotic variance of the naive plug-in estimator.

1. Introduction. In Wicksell (1925) the following stereological problem is studied. Suppose that a number of spheres are embedded in an opaque medium. The item of interest is the distribution function of the sphere radii. Since the medium is opaque, we cannot observe a sample of sphere radii directly. What we can observe is a cross section of the medium, showing circular sections of some spheres. A thorough discussion on the problem and the vast relevant literature can be found in Stoyan, Kendall and Mecke (1987).

If we denote the distribution function of the sphere radii by F_S and assume the centers of the spheres to be distributed according to a homogeneous Poisson process, one can show [see, e.g., Watson (1971)] that the observable circle radii constitute a sample from the density g_C, where

$$g_C(z) = \frac{z}{m_0}\int_z^\infty \frac{dF_S(x)}{\sqrt{x^2 - z^2}}, \tag{1}$$

with $0 < m_0 = \int_0^\infty x\, dF_S(x) < \infty$, the expected sphere radius.

For mathematical convenience, however, we follow Hall and Smith (1988) in considering *squared* radii of both the balls and the circles instead of the radii themselves. In that case the relation between the density g of the observable squared circle radii and the distribution function F of the squared sphere radii follows from (1) and is given by

$$g(z) = \frac{1}{2m_0}\int_{(z,\infty)} \frac{dF(x)}{\sqrt{x - z}},$$

Received May 1993; revised November 1994.
AMS 1991 *subject classifications*. Primary 62G05; secondary 62E20.
Key words and phrases. Inverse problems, minimax rate, asymptotic distribution, arg max functionals.

where $0 < m_0 = \int \sqrt{x}\, dF(x) < \infty$. This integral equation can be inverted, giving an expression of F in terms of g:

$$F(x) = 1 - \frac{2m_0}{\pi} \int_x^\infty \frac{g(z)}{\sqrt{z-x}}\, dz = 1 - \frac{\int_x^\infty (z-x)^{-1/2} g(z)\, dz}{\int_0^\infty z^{-1/2} g(z)\, dz}.$$

Writing

$$V(x) = \int_x^\infty \frac{g(z)}{\sqrt{z-x}}\, dz,$$

we see that

$$F(x) = 1 - \frac{V(x)}{V(0)}.$$

Therefore, the problem of estimating F at a fixed point $x_0 > 0$ is equivalent to the problem of estimating V at two fixed points, that is, x_0 and 0. Taking a closer look at V, we see that V is right-continuous and decreasing on $[0, \infty)$. Moreover, from the relation

$$g(z) = -\frac{1}{\pi} \int_z^\infty \frac{dV(x)}{\sqrt{x-z}}$$

and the fact that g should integrate to 1, it follows that V satisfies

$$\int_0^\infty \sqrt{x}\, dV(x) = -\frac{\pi}{2}.$$

Finally, the requirement that $m_0 \in (0, \infty)$ forces V to be bounded and vanish at infinity. In other words, V must belong to the class $\mathscr{V}$, where

$$\mathscr{V} = \left\{ V\colon V \text{ is a decreasing bounded right-continuous function on } [0,\infty) \text{ with } \lim_{x\to\infty} V(x) = 0 \text{ and } \int_0^\infty \sqrt{x}\, dV(x) = -\frac{\pi}{2} \right\}.$$

Apart from its intrinsic interest, the Wicksell problem has several interesting features as a prototype of an inverse problem. First, it was already clear from, for example, Hall and Smith (1988) that the pointwise rate of convergence of "good" estimators of $V(0)$ should be faster than $n^{-\alpha}$ for $\alpha < \frac{1}{2}$, but the exact rate was not known. In Section 2 we show that, under general conditions, the rate for estimating V at a fixed point cannot be faster than $n^{-1/2}\sqrt{\log n}$.

Second, naive plug-in estimators of the distribution function exist which have the right pointwise rate of converge, but look very strange (to say the least) as estimators of a distribution function, since they have infinite discontinuities at a set of points becoming dense in the support of the distribution function, as the sample size tends to infinity (for a picture, see Section 5). These estimators violate both the monotonicity and the range constraints for distribution functions. We discuss this in Section 3.

In Section 4 we define an isotonized naive estimator $\hat{V}_n$ of V. It turns out that the isotonic estimator of V beats the naive estimator on all fronts: it does not have the undesirable properties of the naive estimator, and the asymptotic variance of the isotonized naive estimator is exactly half the asymptotic variance of the naive estimator. From the monotone estimator of V we get an estimator $\hat{F}_n$ of F, which is actually a distribution function, and defined by

$$\hat{F}_n(x) = 1 - \hat{V}_n(x)/\hat{V}_n(0).$$

The asymptotic variance of $\hat{F}_n(x)$ is also exactly one-half times the asymptotic variance of the corresponding naive estimator.

The isotonic estimator $\hat{V}_n$ of V exhibits some other interesting features. As usual in this type of analysis, the limiting distribution function is given by the distribution of the location of the maximum of a Gaussian process minus a parabola. However, in this case the Gaussian process is degenerate, with the result that the limiting distribution is in fact just a normal distribution.

Furthermore, in order to obtain the limit distribution of $\hat{F}_n$, we need to study the limiting behavior of the derivative of the convex envelope of a certain empirical process at zero. In general, the behavior at zero of estimators obtained in this way is somewhat pathological. As an example, the maximum likelihood (Grenander) estimator of a decreasing density at zero, which is just the right derivative of the concave majorant of the empirical distribution function at zero, is an inconsistent estimator of the real underlying density at zero [see Woodroofe and Sun (1993), who also provide a remedy]. In our case, this does not turn out to be the case: the derivative of the support function at zero is in fact a consistent estimator of the quantity of interest.

As some other remarkable features of the Wicksell problem we note that smoothness assumptions do not seem to be helpful for getting faster rates in pointwise estimation of the distribution function and also that the $n^{-1/2}$-rate in the estimation of certain "smooth functionals," like the mean, seems in general to be unattainable. The latter is in contrast with deconvolution problems with which the Wicksell problem is related. In general, if the observations consist of a sample of random variables $Z_i = X_i + Y_i$, where the Y_i have a known distribution and the interest is in estimating the distribution function of the X_i's, the mean of the X_i's can be estimated at $n^{-1/2}$-rate, but the pointwise rate is much worse [see, e.g., Groeneboom and Wellner (1992)]. In comparison, for the Wicksell problem the pointwise rate is quite good $(n^{-1/2}\sqrt{\log n})$, but there is no improvement for the estimation of moments. Smoothness assumptions on the underlying distribution function do not seem to help here, the difficulties really seem to derive from the nature of the inverse problem, in particular, from the extra parameter m_0. However, *ratios* of moments can indeed be estimated at the $n^{-1/2}$ rate (in that case we lose the extra parameter in the division).

In this paper we will not consider the nonparametric maximum likelihood estimator of the distribution function in Wicksell's problem. For the definition and some properties of that estimator we refer to van Es (1991) and Jongbloed (1991).

2. Minimax lower bound on the optimal rate of convergence. In this section we will prove that if one does not want to assume too much concerning V as introduced in Section 1, one cannot estimate $V(x_0)$ at a rate faster than $n^{-1/2}\sqrt{\log n}$, where n is the sample size of the observed circle radii. It will then follow that one also cannot estimate $F(x_0)$ at a rate faster than $n^{-1/2}\sqrt{\log n}$, for any point $x_0 > 0$ in the interior of the support of F.

For even if $V(0)$ is known, one cannot estimate $V(x_0)$ at a rate faster than $n^{-1/2}\sqrt{\log n}$ [basically because, as shown below, the estimation problem is a local one, and knowledge of $V(x_1)$ at a point $x_1 \neq x_0$ does not help to improve the rate in estimating $V(x_0)$], whereas a faster rate of convergence in estimating $F(x_0)$, under the assumption that $V(0)$ is known, would lead to a faster rate of convergence in estimating $V(x_0)$, since

$$V(x_0) = V(0)(1 - F(x_0)).$$

Fix $x_0 \geq 0$ and let Θ be defined by

$$\Theta = \left\{V \in \mathscr{V}: \frac{dV}{dx}(x) < 0 \text{ for } x \in (x_0 - \delta, x_0 + \delta) \text{ for some } \delta < 0\right\}$$

if $x_0 > 0$, and by

$$\Theta = \left\{V \in \mathscr{V}: \frac{dV}{dx}(x) < 0 \text{ for } x \in (0, \delta) \text{ for some } \delta > 0\right\}$$

if $x_0 = 0$. Let the minimax risk $R(n)$ for the class Θ be defined as follows:

$$R(n) = \inf_{T_n} \sup_{V \in \Theta} E_V |V(x_0) - T_n|,$$

where the infimum is taken over all estimators T_n, based on a sample from the density g, where

$$g(z) = -\frac{1}{\pi}\int_z^\infty \frac{dV(x)}{\sqrt{x - z}}.$$

We even allow T_n to depend on the value of V at a fixed point $x_1 \neq x_0$. The following theorem shows that $R(n)$ cannot vanish at a rate faster than $n^{-1/2}\sqrt{\log n}$.

THEOREM 1. *We have*

$$\liminf_{n \to \infty} \sqrt{\frac{n}{\log n}}\, R(n) > 0.$$

PROOF. Note that for each subclass Θ_n of Θ the following inequality trivially holds:

$$R(n) \geq \inf_{T_n} \sup_{V \in \Theta_n} E_V |V(x_0) - T_n|. \tag{2}$$

For each n we will construct a subset Θ_n, containing just a fixed $V_0 \in \Theta$ and some perturbed version of V_0, and find a minimax lower bound for the discrimination problem for these two functions. First we consider the case $x_0 > 0$.

Define for $\varepsilon > 0$ and $\delta > 0$ such that the derivative of V_0 is strictly negative in the region $[x_0 - 2\delta, x_0 + 2\delta]$, the function $\phi_\varepsilon^{(1)}$ as follows:

$$\phi_\varepsilon^{(1)}(x) = \frac{\varepsilon}{\sqrt{\log(1/\varepsilon)}} \frac{\sin^2((x - x_0)/2\varepsilon)}{x - x_0} 1_{[x_0 - \delta, x_0 + \delta]}(x).$$

Based on $\phi_\varepsilon^{(1)}$ we can define the function $\phi_\varepsilon^{(2)}$:

$$\phi_\varepsilon^{(2)}(x) = v_\varepsilon\big(1_{[x_0 - \delta, x_0)}(x) - 1_{[x_0, x_0 + \delta)}(x)\big),$$

where

$$v_\varepsilon = \frac{3\int_{[x_0 - \delta, x_0 + \delta]} \sqrt{x}\, \phi_\varepsilon^{(1)}(x)\, dx}{2\big(2x_0^{3/2} - (x_0 + \delta)^{3/2} - (x_0 - \delta)^{3/2}\big)}.$$

Using these functions we define a perturbed version V_ε of V_0 in the following way:

$$V_\varepsilon(x) = V_0(x) + \int_0^x \big(\phi_\varepsilon^{(1)}(y) - \phi_\varepsilon^{(2)}(y)\big)\, dy.$$

Since

$$\sup_{x \in \mathbb{R}} \phi_\varepsilon^{(2)}(x) = v_\varepsilon \asymp \frac{\varepsilon}{\sqrt{\log(1/\varepsilon)}} \quad \text{and} \quad \sup_{x \in \mathbb{R}} \phi_\varepsilon^{(1)}(x) \asymp \frac{1}{\sqrt{\log(1/\varepsilon)}}$$

for $\varepsilon \downarrow 0$, where the relation $\asymp$ means that the ratio of the left-hand and right-hand sides stays away from zero and ∞, the function V_ε is, for ε sufficiently small, a decreasing function on $[0, \infty)$. Moreover,

$$\int_0^\infty \sqrt{x}\, dV_\varepsilon(x) = -\frac{\pi}{2}$$

and, for ε sufficiently small, V_ε has a derivative which is strictly negative in the region $[x_0 - 2\delta, x_0 + 2\delta]$. Hence, for ε sufficiently small, $V_\varepsilon \in \Theta$. For a vanishing sequence of positive numbers (ε_n), to be specified below, we define

$$\Theta_n = \{V_0, V_{\varepsilon_n}\}.$$

The next step is to bound the right-hand side of (2) from below by an expression involving distances between Wicksell densities g corresponding to the members of Θ_n. For this, we need the following notions.

Let P and Q be the probability measured on $\mathbb{R}^k$, with derivatives p and q, respectively, with respect to Lebesque measure l. The *inf-measure* corresponding to P and Q is defined by its density:

$$\frac{d(P \wedge Q)}{dl}(x) = \min(p(x), q(x)), \qquad x \in \mathbb{R}^k.$$

The *squared Hellinger distance* between P and Q is defined as

$$H^2(P, Q) = \frac{1}{2}\int_{\mathbb{R}^k}\left(\sqrt{p(x)} - \sqrt{q(x)}\right)^2 dl(x).$$

The squared Hellinger distance has the property that, writing $P^{\otimes n}$ for the n-fold product measure of a probability measure P on the real line,

$$1 - H^2(P^{\otimes n}, Q^{\otimes n}) = \left(1 - H^2(P, Q)\right)^n.$$

Using the triangle inequality and Le Cam's inequality [Le Cam (1973)], we get, for any estimator T_n, which may depend on $V_0(x_1)$ for some $x_1 \neq x_0$ and n sufficiently large,

$$\begin{aligned}\inf_{T_n}\sup_{V\in\Theta_n} E_V|V(x_0) - T_n| &\geq \frac{1}{2}\left(V_{\varepsilon_n}(x_0) - V_0(x_0)\right)\|P_{\varepsilon_n}^{\otimes n} \wedge P_0^{\otimes n}\|_1\\ &\geq \frac{1}{4}\left(V_{\varepsilon_n}(x_0) - V_0(x_0)\right)\left(1 - H^2(P_0, P_{\varepsilon_n})\right)^{2n},\end{aligned}$$

where P_0 and P_{ε_n} denote the probability measures corresponding to the densities g_0 and g_{ε_n} (see the last formula before Theorem 1), respectively. Clearly,

$$|V_{\varepsilon_n}(x_0) - V_0(x_0)| \asymp \varepsilon_n\sqrt{\log 1/\varepsilon_n}\,.$$

On the other hand,

$$\begin{aligned}H^2(P_0, P_{\varepsilon_n}) &= \frac{1}{2}\int_{\mathbb{R}}\left(\sqrt{g_0(x)} - \sqrt{g_{\varepsilon_n}(x)}\right)^2 dx\\ &= \frac{1}{2}\int_{\mathbb{R}} g_0(x)\left(1 - \sqrt{1 + \frac{g_{\varepsilon_n}(x) - g_0(x)}{g_0(x)}}\right)^2 dx\\ &\leq \frac{1}{4}\int_{[0,\infty)}\frac{\left(g_{\varepsilon_n}(z) - g_0(z)\right)^2}{g_0(z)}\,dz.\end{aligned}$$

However, since $g_{\varepsilon_n}(z) = g_0(z)$ for $z \geq x_0 + \delta$ and since $V \in \Theta$ implies that g_0 stays away from zero on $[0, x_0 + \delta]$, we immediately obtain that

$$H^2(P_0, P_{\varepsilon_n}) \leq c\|g_{\varepsilon_n} - g_0\|_2^2,$$

for some positive constant c.

To complete the proof we need the following lemma, to be proved in the Appendix.

LEMMA 1. *For $n \to \infty$ we have*

$$\|g_{\varepsilon_n} - g_0\|_2 \asymp \varepsilon_n.$$

Substituting $1/\sqrt{n}$ for ε_n throughout, we obtain

$$\liminf_{n\to\infty} \sqrt{\frac{n}{\log n}}\, R(n) \geq c,$$

for some positive constant c, proving the result for the case $x_0 > 0$.

We still have to prove the result for the case $x_0 = 0$. To this end we redefine $\phi_\varepsilon^{(1)}$ in the following way:

$$\phi_\varepsilon^{(1)}(x) = \frac{\varepsilon}{\sqrt{\log(1/\varepsilon)}} \frac{\sin^2(x/2\varepsilon)}{x} 1_{(0,\delta]}(x).$$

We also define the constants

$$\alpha_\varepsilon = \int_0^\delta \sqrt{x}\, \phi_\varepsilon^{(1)}(x)\, dx \sim \frac{\varepsilon}{\sqrt{\log 1/\varepsilon}}$$

and

$$\beta_\varepsilon = \int_0^\delta \phi_\varepsilon^{(1)}(x)\, dx \sim \varepsilon\sqrt{\log 1/\varepsilon}$$

and the function $\phi_\varepsilon^{(2)}$ by

$$\phi_\varepsilon^{(2)}(x) = \tfrac{3}{2}\alpha_\varepsilon \delta^{-3/2} 1_{[0,\delta]}(x).$$

Our perturbed version of V_0 now becomes

$$V_\varepsilon(x) = V_0(x) - \beta_\varepsilon + \int_0^x \left(\phi_\varepsilon^{(1)}(y) - \phi_\varepsilon^{(2)}(y)\right) dy + \tfrac{3}{2}\alpha_\varepsilon \delta^{-1/2}.$$

for $\varepsilon > 0$ sufficiently small we again have that $V_\varepsilon \in \Theta$. The remaining part of the proof proceeds along the same lines as the proof for the case $x_0 > 0$. □

3. Asymptotics of a naive estimator. Suppose we have an estimator G_n (which is a distribution function) for the distribution function G of the squared circle radii. Our *plug-in* estimator $V_n(x)$ of $V(x)$ is then defined as follows:

$$V_n(x) = \int_x^\infty \frac{dG_n(z)}{\sqrt{z - x}}.$$

If the estimator G_n is constructed without using the special structure of the problem, we call these plug-in estimators *naive* estimators. The naive estimator most frequently used is the one where G_n is the empirical distribution function of the squared circle radii. For finite n, however, V_n is *increasing* between successive observation points and has infinite jumps at the observation points, implying $\|V_n - V\|_\infty = \infty$ for all n. Since the function V is bounded and decreasing, the function V_n is a strange estimate for V indeed.

The asymptotics of the naive estimator based on the empirical distribution function are given in the theorem below.

THEOREM 2. *The estimator V_n is pointwise strongly consistent, that is, for all $x \geq 0$ the following holds:*

$$V_n(x) \to V(x) \quad a.s. \text{ for } n \to \infty.$$

For each $x \geq 0$ where $g(x) < \infty$, we have

$$\sqrt{\frac{n}{\log n}}\,(V_n(x) - V(x)) \to_{\mathscr{D}} \mathscr{N}(0, g(x)) \quad \text{as } n \to \infty.$$

PROOF. The first assertion follows from the strong law of large numbers. Applying the central limit theorem for sums of independent random variables with infinite variance [see, e.g., Chow and Teicher (1988), Theorem 4, page 305] to the random variables $1_{(x,\infty)}(Z_i)/\sqrt{Z_i - x}$, the second part of the theorem is established. □

From Theorem 2 we get the following corollary.

COROLLARY 1. *Let F be the distribution function of the squared sphere radii. Then we have for each $x > 0$ such that $g(x) < \infty$,*

$$\sqrt{\frac{n}{\log n}}\left\{1 - \frac{V_n(x)}{V_n(0)} - F(x)\right\} \to_{\mathscr{D}} \mathscr{N}\left(0, \frac{g(x)V(0)^2 + g(0)V(x)^2}{V(0)^4}\right) \quad \text{as } n \to \infty.$$

PROOF. Using $1 - F(x) = V(x)/V(0)$, we can write

$$\begin{aligned} 1 - \frac{V_n(x)}{V_n(0)} - F(x) &= \frac{V(x)}{V(0)} - \frac{V_n(x)}{V_n(0)} \\ &= \frac{V(x) - V_n(x)}{V_n(0)} + \frac{V(x)\{V_n(0) - V(0)\}}{V(0)V_n(0)}. \end{aligned}$$

The corollary now follows from Theorem 2, Slutsky's theorem and the asymptotic independence of $\sqrt{n}(\log n)^{-1/2}(V_n(x) - V(x))$ and $\sqrt{n}(\log n)^{1/2}(V_n(0) - V(0))$. □

4. Isotonized naive estimators. Based on a naive estimator, which is in general *not* a decreasing bounded function, we can define an estimator which is a bounded decreasing function. We simply define as our estimator that function within the class of bounded decreasing right-continuous functions which has the smallest L_2-distance to the naive estimator. Before we proceed with the estimation problem, we state the following general lemma, giving an explicit graphical interpretation of the implicitly defined projection estimate. The proof of the lemma can be found in the Appendix.

LEMMA 2. *Let ϕ be an a.e. continuous nonnegative integrable function on $[0, \infty)$ with compact support. Define the function Φ as follows*:

$$\Phi(x) = \int_0^x \phi(y)\, dy \quad \text{for } x \geq 0.$$

Let Φ^* be the concave majorant of Φ. If we define

$$\phi^*(x) = \frac{d^{(r.h.s.)}\Phi^*}{dx}(x),$$

the right derivative of Φ^*, we have that

$$\int_0^\infty (\phi(x) - \psi(x))^2\, dx \geq \int_0^\infty (\phi(x) - \phi^*(x))^2\, dx + \int_0^\infty (\phi^*(x) - \psi(x))^2\, dx,$$

for all functions $\psi \in \mathscr{F}$, where

$$\mathscr{F} = \{\psi: [0,\infty) \to [0,\infty): \psi \text{ is decreasing and right-continuous}\}.$$

We can apply this lemma to the function $\phi = V_n$. If we define U_n as

$$U_n(x) = \int_0^x V_n(y)\, dy = 2\int_0^\infty \sqrt{z}\, dG_n(z) - 2\int_x^\infty \sqrt{z-x}\, dG_n(z)$$

and define U_n^* to be the concave majorant of U_n, our estimator of $V(x)$ is $U_n^{*,r}(x)$, the right derivative of U_n^* at x. We refer to estimators $U_n^{*,r}$ as *isotonized naive estimators*. For isotonized naive estimators we have the following theorem.

THEOREM 3. *Suppose that the distribution function F of the squared sphere radii satisfies the following*:

(i) $$\int_0^\infty y\, dF(y) < \infty;$$

(ii) *the initial estimator G_n of G has the property that, for each $x \geq 0$,*

$$\int_x^\infty \sqrt{z-x}\, dG_n(z) \to \int_x^\infty \sqrt{z-x}\, g(z)\, dz,$$

a.s. for $n \to \infty$.

Then we have with probability 1, *for each $x > 0$,*

$$V(x-) \geq \limsup_{n\to\infty} U_n^{*,l}(x) \geq \liminf_{n\to\infty} U_n^{*,r}(x) \geq V(x)$$

and

$$\liminf_{n\to\infty} U_n^{*,r}(0) \geq V(0).$$

PROOF. First of all note that

$$\int_x^\infty \sqrt{z-x}\,g(z)\,dz = \frac{\pi}{4m_0}\left\{\int_0^\infty y\,dF(y) - \int_0^x (1-F(y))\,dy\right\}.$$

If we define

$$U(x) = \frac{\pi}{2m_0}\int_0^x (1-F(y))\,dy \qquad \left(= \int_0^x V(y)\,dy\right),$$

we have, by assumption (ii),

$$U_n(x) \to U(x) \quad \text{a.s.}$$

Since the functions U_n are nondecreasing, U is a bounded [by (i)] continuous nondecreasing function, and since

$$\lim_{x\to\infty} U_n(x) = 2\int_0^\infty \sqrt{z}\,dG_n(z) \to 2\int_0^\infty \sqrt{z}\,dG(z) = \lim_{x\to\infty} U(x) \quad \text{a.s.},$$

we have

$$\|U_n - U\|_\infty \to 0 \quad \text{a.s.} \tag{3}$$

By Marshall's lemma [Robertson, Wright and Dykstra (1988), page 329], it follows that, with probability 1, $\|U_n^* - U\|_\infty \to 0$, whereas Lemma 7.2.1 in Robertson, Wright and Dykstra (1988) yields

$$U^l(x) \ge \limsup_{n\to\infty} U_n^{*,l}(x) \ge \liminf_{n\to\infty} U_n^{*,r}(x) \ge U^r(x) \quad \text{a.s.},$$

for each $x > 0$ and

$$\liminf_{n\to\infty} U_n^{*,r}(0) \ge V(0) \quad \text{a.s.}$$

Observing that

$$U^l(x) = V(x-) \quad \text{and} \quad U^r(x) = V(x),$$

the results follows. □

Note that, if F is continuous, Theorem 3 implies that, with probability 1, for each $\varepsilon > 0$, $U_n^{*,r} \to V$ uniformly on $[\varepsilon, \infty)$.

The behavior of $U_n^{*,r}(0)$ and asymptotic distribution theory for $U_n^{*,r}(x)$ are hard to study in the general setting, that is, for arbitrary initial estimators G_n of G. Therefore we now consider the case that G_n is the empirical distribution function of $Z_1,\dots,Z_n$, the sample of observable squared profile radii. In that case we write $\hat{V}_n$ for $U_n^{*,r}$ and we have

$$U_n(x) = \frac{2}{n}\left\{\sum_{i=1}^{n}\sqrt{Z_{(i)}} - \sum_{i=i^*+1}^{n}\sqrt{Z_{(i)}-x}\right\},$$

where

$$i^* = \max\{i: Z_{(i)} < x\}.$$

By the strong law of large numbers, which can be applied if condition (i) of Theorem 3 is satisfied, it is easily seen that (ii) in Theorem 3 is then also automatically satisfied. Because U_n is not only nondecreasing, but also convex between successive observation points, we have that the concave majorant of U_n equals the concave majorant of the diagram consisting of the points $(Z_{(i)}, U_n(Z_{(i)}))$ for $0 \le i \le n$. Therefore, algorithms from the theory of isotonic regression can be applied to compute $\hat{V}_n$ in practical situations.

Woodroofe and Sun (1993) proved that the maximum likelihood or Grenander estimator of a decreasing density on $[0, \infty)$ is inconsistent at zero, whereas it is strongly consistent for all $x > 0$. This Grenander estimator can, for $x > 0$, be viewed as the left derivative of the concave majorant of the empirical distribution function of the data, and its value at zero is determined by right-continuity of the estimator at zero. Woodroofe and Sun showed that, asymptotically, the value of the Grenander estimator at zero is always too large. In our case, where we consider a derivative of the concave majorant of an empirical process which is not the empirical distribution function, we have already seen in Theorem 3 that $\hat{V}_n(0)$ is almost surely not too small asymptotically. The next theorem shows that our U_n-process is better behaved than the empirical distribution function based on a sample from a decreasing density, in the sense that the random variable $\hat{V}_n(0)$ is weakly consistent.

THEOREM 4. *Let condition* (i) *of Theorem* 3 *be satisfied, and let* G_n *be the empirical distribution function of a sample of size n from the distribution function G of the squared circle radii. Then*

$$\hat{V}_n(0) \to V(0) \quad \textit{in probability, as } n \to \infty.$$

PROOF. From Theorem 3 we know that, with probability 1,

$$\liminf_{n \to \infty} \hat{V}_n(0) \ge V(0).$$

Therefore if suffices to show that, for each $\varepsilon > 0$ and $\eta > 0$,

$$\mathbb{P}\{\hat{V}_n(0) - V(0) > \varepsilon\} < \eta \quad \text{for } n \text{ large enough.}$$

Fix $\varepsilon > 0$, let $R > 0$ be a constant and let $\varepsilon_n = n^{-1/4}$ (other choices of ε_n are also possible, but this choice will do). Now note that

$$\begin{aligned}
\mathbb{P}\{\hat{V}_n(0) - V(0) > \varepsilon\} &= \mathbb{P}\left\{\exists\, t > 0\colon \frac{U_n(t)}{t} > V(0) + \varepsilon\right\} \\
&= \mathbb{P}\{\exists\, t > 0\colon U_n(t) > t(V(0) + \varepsilon)\} \\
&\le \mathbb{P}\{\exists\, t \in (0, \varepsilon_n]\colon U_n(t) > t(V(0) + \varepsilon)\} \\
&\quad + \mathbb{P}\{\exists\, t \in (\varepsilon_n, R]\colon U_n(t) > t(V(0) + \varepsilon)\} \\
&\quad + \mathbb{P}\{\exists\, t > R\colon U_n(t) > t(V(0) + \varepsilon)\} \\
&=: p_1(n) + p_2(n) + p_3(n).
\end{aligned}$$

For $p_3(n)$ we can write

$$\begin{aligned} p_3(n) &= \mathbb{P}\{\exists\, t > R: U_n(t) > t(V(0) + \varepsilon)\} \\ &\le \mathbb{P}\{\exists\, t > R: U_n(t) - U(t) > \varepsilon R\} \\ &\le \mathbb{P}\{\|U_n - U\|_\infty > \varepsilon R\}, \end{aligned}$$

where we use that, for all $z \ge 0$,

$$U(z) \le V(0) z. \tag{4}$$

Since, with probability 1, $\|U_n - U\|_\infty \to 0$ [see (3)], it follows that $p_3(n) < \eta/3$ for all n large enough.

We will now consider the term $p_2(n)$:

$$\begin{aligned} p_2(n) &= \mathbb{P}\{\exists\, t \in (\varepsilon_n, R]: U_n(t) > t(V(0) + \varepsilon)\} \\ &\le \mathbb{P}\Big\{\max_{\varepsilon_n \le t \le R} \sqrt{n}\,(U_n(t) - U(t)) > \sqrt{n}\,((V(0) + \varepsilon)\varepsilon_n - U(\varepsilon_n))\Big\} \\ &\le \mathbb{P}\Big\{\max_{0 \le t \le R} \sqrt{n}\,(U_n(t) - U(t)) > \varepsilon\sqrt{n}\,\varepsilon_n\Big\}, \end{aligned}$$

where, in the last step, we use (4). Note that $\sqrt{n}\,(U_n - U)$ converges in distribution in $C[0, \infty)$, with the topology of uniform convergence on compacta, to a Gaussian process W. This implies that there exists a number $M > 0$ such that

$$\mathbb{P}\Big\{\sup_{t \in [0, R]} |W(t)| > M\Big\} < \frac{\eta}{6}.$$

Therefore, since we chose $\varepsilon_n = n^{-1/4}$, we obtain, for $n \ge (M/\varepsilon)^4$,

$$\begin{aligned} p_2(n) &\le \mathbb{P}\Big\{\max_{0 \le t \le R} \sqrt{n}\,(U_n(t) - U(t)) > \varepsilon n^{1/4}\Big\} \\ &\le \mathbb{P}\Big\{\max_{0 \le t \le R} \sqrt{n}\,(U_n(t) - U(t)) > M\Big\}. \end{aligned}$$

The continuous mapping theorem finally yields that $p_2(n) \le \eta/3$ for all n sufficiently large.

The argument below shows that $p_1(n) < \eta/3$ for all n sufficiently large. Since U_n is convex between successive observations Z_i, we can write

$$\begin{aligned} p_1(n) &= \mathbb{P}\{\exists\, 1 \le i \le n: U_n(Z_i) > (V(0) + \varepsilon) Z_i \text{ and } Z_i \in (0, \varepsilon_n]\} \\ &\le n\mathbb{P}\{U_n(Z_1) > (V(0) + \varepsilon) Z_1 \text{ and } Z_1 \in (0, \varepsilon_n]\} \\ &= n\int_0^{\varepsilon_n} \mathbb{P}\{U_n(z) > (V(0) + \varepsilon) z \mid Z_1 = z\} g(z)\, dz \\ &= n\int_0^{\varepsilon_n} \phi_n(z) g(z)\, dz, \end{aligned}$$

where

$$\phi_n(z) = \mathbb{P}\left\{\frac{2}{n}\sum_{i=2}^{n}\left(\sqrt{Z_i} - \sqrt{Z_i - z}\,1_{[z,\infty)}(Z_i)\right) + \frac{2}{n}\sqrt{z} > (V(0)+\varepsilon)z \,\middle|\, Z_1 = z\right\}.$$

We will now estimate ϕ_n from above on the interval $(0, \varepsilon_n]$. For that we use that, for $z \geq 16\varepsilon^{-2}n^{-2}$, we have that $\varepsilon z - 2\sqrt{z}\,n^{-1} \geq \varepsilon z/2$. Splitting the interval $(0, \varepsilon_n]$ as

$$(0, \varepsilon_n] = (0, 16\varepsilon^{-2}n^{-2}] \cup (16\varepsilon^{-2}n^{-2}, \varepsilon_n]$$

[note that to this end we must take n is sufficiently large, i.e., $n \geq (16\varepsilon^{-2})^{4/7}$, since $\varepsilon_n = n^{-1/4}$], we can write

$$\phi_n(z) \leq \begin{cases} 1, & \text{for } z \in (0, 16\varepsilon^{-2}n^{-2}], \\ \mathbb{P}\left\{\frac{2}{n}\sum_{i=2}^{n}\left(\sqrt{Z_i} - \sqrt{Z_i - z}\,1_{[z,\infty)}(Z_i)\right) > \left(V(0) + \frac{\varepsilon}{2}\right)z\right\}, & \\ & \text{for } z \in (16\varepsilon^{-2}n^{-2}, \varepsilon_n]. \end{cases}$$

We will now concentrate on the nontrivial part of the upper bound for ϕ_n. Redefine a random variable Z_1 with density g which is independent of the old Z_1 and of $Z_2, Z_3, \ldots$. We may write

$$\begin{aligned} &\mathbb{P}\left\{\frac{2}{n}\sum_{i=2}^{n}\left(\sqrt{Z_i} - \sqrt{Z_i - z}\,1_{[z,\infty)}(Z_i)\right) > \left(V(0) + \frac{\varepsilon}{2}\right)z\right\} \\ &\quad\leq \mathbb{P}\left\{\frac{2}{n}\sum_{i=1}^{n}\left(\sqrt{Z_i} - \sqrt{Z_i - z}\,1_{[z,\infty)}(Z_i)\right) > \left(V(0) + \frac{\varepsilon}{2}\right)z\right\} \\ &\quad\leq \left(V(0)z + \frac{\varepsilon z}{2} - U(z)\right)^{-2}\operatorname{Var}(U_n(z)) \leq \frac{4}{\varepsilon^2 z^2}\operatorname{Var}(U_n(z)), \end{aligned}$$

where we use Chebyshev's inequality and (4). However, since $U_n(z)$ is the mean of the n independent identically distributed random variables, we have

$$n\operatorname{Var}(U_n(z)) = 4\operatorname{Var}\left(\sqrt{Z_1} - \sqrt{Z_1 - z}\,1_{[z,\infty)}(Z_1)\right),$$

which does not depend on n. Writing out this variance, we obtain

$$\begin{aligned} &\operatorname{Var}\left(\sqrt{Z_1} - \sqrt{Z_1 - z}\,1_{[z,\infty)}(Z_1)\right) \\ &\quad= \int_0^z y g(y)\,dy - 2\int_z^\infty y\sqrt{1 - \frac{z}{y}}\,g(y)\,dy + \int_z^\infty (2y - z)g(y)\,dy - U(z)^2. \end{aligned}$$

The first term can be bounded from above as follows:

$$\int_0^z y g(y)\,dy = \int_0^z y^{3/2}y^{-1/2}g(y)\,dy \leq z^{3/2}V(0). \tag{5}$$

Since $U(z)^2 = O(z^2)$ for $z \downarrow 0$, we obtain for $z \downarrow 0$ that

$$\begin{aligned}
&\operatorname{Var}\left(\sqrt{Z_1} - \sqrt{Z_1 - z}\, 1_{[z,\infty)}(Z_1)\right)\\
&\quad = 2\int_z^\infty \frac{z^2}{8y} g(y)\, dy\\
&\qquad - 2\int_z^\infty y\left(\sqrt{1 - \frac{z}{y}} - 1 + \frac{z}{2y} + \frac{z^2}{8y^2}\right) g(y)\, dy + O(z^{3/2}).
\end{aligned}$$

Using approximations similar to (5), it can be shown that the first term is of order $O(z^{3/2})$ for $z \downarrow 0$. From Taylor's theorem it follows that the second term is of the same order for $z \downarrow 0$. Note that no local smoothness properties of g are needed. The only assumption on the unknown distribution is that $V(0) < \infty$.

Therefore, there exists a positive constant c such that, for each z below some positive number ν,

$$\mathbb{P}\left\{\frac{2}{n}\sum_{i=2}^{n}\left(\sqrt{Z_i} - \sqrt{Z_i - z}\, 1_{[z,\infty)}(Z_i)\right) > \left(V(0) + \frac{\varepsilon}{2}\right)z\right\} \le \frac{cz^{-1/2}}{n\varepsilon^2}.$$

Consequently, for n large enough (such that $16\varepsilon^{-2}n^{-2} \le \varepsilon_n < \nu$), we may write

$$\begin{aligned}
p_1(n) &\le n\int_0^{\varepsilon_n} \phi_n(z) g(z)\, dz\\
&\le n\int_0^{16/(\varepsilon^2 n^2)} g(z)\, dz + n\int_{16/(\varepsilon^2 n^2)}^{\varepsilon_n} \frac{c}{n\varepsilon^2} z^{-1/2} g(z)\, dz\\
&\le \left(\frac{4}{\varepsilon} + \frac{c}{\varepsilon^2}\right)\int_0^{\varepsilon_n} z^{-1/2} g(z)\, dz.
\end{aligned}$$

By the dominated convergence theorem we therefore have that $p_1(n) \to 0$ for $n \to \infty$. This shows that $p_1(n) < \eta/3$ eventually, implying that

$$\mathbb{P}\left\{\hat{V}_n(0) - V(0) > \varepsilon\right\} < \eta \quad \text{eventually.} \qquad \square$$

From Theorems 3 and 4 it follows that, for each fixed $x \ge 0$, $\hat{V}_n(x)$ converges to $V(x)$ in probability. In the theorem below the asymptotic distribution of $\hat{V}_n(x)$ is given. It is interesting to see this result in connection with Theorem 2, where the asymptotic distribution of the naive estimator V_n based on the empirical distribution function is given. It turns out that the rates of convergence of both $V_n(x)$ and $\hat{V}_n(x)$ are $n^{-1/2}\sqrt{\log n}$, which is optimal in the sense of Section 2. A surprising difference, however, occurs in the asymptotic variances of $V_n(x)$ and $\hat{V}_n(x)$. The asymptotic variance of $\hat{V}_n(x)$ is exactly half the asymptotic variance of $V_n(x)$.

THEOREM 5. *Let $x \ge 0$, and suppose that F has a density f which is strictly positive at x and continuous in a neighborhood of x (if $x = 0$, we mean a right* *neighborhood). Then we have*

$$\sqrt{\frac{n}{\log n}}\left(\hat{V}_n(x) - V(x)\right) \to_{\mathscr{D}} \mathscr{N}\left(0, \frac{1}{2}g(x)\right) \quad \textit{as } n \to \infty.$$

PROOF. The proof essentially consists of three parts. The first part is to write the event

$$\left\{\sqrt{\frac{n}{\log n}}\left(\hat{V}_n(x) - V(x)\right) \le a\right\}$$

in terms of the location of the maximum (arg max) of a stochastic process $t \mapsto \tilde{Z}_n(t) - at$ [see (7)]. The second step is to show convergence in distribution of the sequence $(\tilde{Z}_n)$ to a process $t \mapsto Z(t)$ on the space $(C(-\infty, \infty), d)$, where d is the metric of uniform convergence on compacta. This result is established by showing the asserted convergence for a sequence (Z_n) of processes, where $d(Z_n, \tilde{Z}_n) \to 0$. The final step is to prove that from the convergence in distribution of $(\tilde{Z}_n(t) - at)$ to $Z(t) - at$ it follows that the arg max of $(\tilde{Z}_n(t) - at)$ converges in distribution to the arg max of $Z(t) - at$. Since the arg max functional is certainly not continuous on $(C(-\infty, \infty), d)$, the simplest form of the continuous mapping theorem cannot be applied. Some extra work has to be done. From Lemma 4 it follows that the sequence of locations of the maximum of $(\tilde{Z}_n(t) - at)$ is uniformly tight. This enables us to localize the problem: uniformly in n the probability that the arg $\max_{t\in\mathbb{R}}$ of $(\tilde{Z}_n(t) - at)$ equals the arg $\max_{t\in[-M, M]}$ of the same process can be made arbitrarily close to 1 by choosing M sufficiently large. Moreover, since the limiting process turns out almost surely to have a unique maximum and the arg max functional is, for all $M > 0$, continuous on the subset of $(C(-M, M), \|\cdot\|_\infty)$ consisting of functions having a unique maximum, the asserted convergence in distribution follows.

Fix $x \ge 0$ such that x satisfies the conditions stated above. Define, for $a > 0$, the process T_n as follows:

$$T_n(a) = \inf\{t \ge 0 \colon U_n(t) - at \text{ is maximal}\}.$$

The relation between T_n and $\hat{V}_n$ is given by

$$T_n(a) \le t \iff \hat{V}_n(t) \le a. \tag{6}$$

From this relation it follows immediately that, for each vanishing sequence (δ_n) of positive numbers and for $a > 0$, the following holds:

$$\delta_n^{-1}(V_n(x) - V(x)) \le a \iff \delta_n^{-1}(T_n(a_0 + \delta_n a) - x) \le 0, \tag{7}$$

where we write a_0 for $V(x)$. Note that a_0 is fixed.

It is now convenient to introduce the following notation:

$$I_x = \begin{cases} [0, \infty), & \text{if } x = 0, \\ \mathbb{R}, & \text{if } x > 0. \end{cases}$$

Define the process $\tilde{Z}_n$ in $C(I_x)$, the space of the continuous functions on I_x, as follows:

$$\tilde{Z}_n(t) = \delta_n^{-2}\left[U_n(x+\delta_n t) - U_n(x) - a_0\delta_n t\right].$$

For each fixed $a \in \mathbb{R}$ we can, for all n sufficiently large, write

$$\begin{aligned}
&\delta_n^{-1}(T_n(a_0+\delta_n a) - x)\\
&\quad = \delta_n^{-1}\big(\inf\{x+t \ge 0\colon U_n(x+t) - a_0 x\\
&\qquad\qquad -a_0 t - \delta_n a x - \delta_n a t \text{ is maximal}\} - x\big)\\
&\quad = \delta_n^{-1}\inf\{t \ge -x\colon U_n(x+t) - U_n(x) - a_0 t - \delta_n a t \text{ is maximal}\}\\
&\quad = \inf\{t \ge -\delta_n^{-1}x\colon U_n(x+\delta_n t) - U_n(x) - a_0\delta_n t - \delta_n^2 a t \text{ is maximal}\}\\
&\quad = \inf\{t \ge -\delta_n^{-1}x\colon \tilde{Z}_n(t) - at \text{ is maximal}\}.
\end{aligned}$$

What is used essentially is that the location of a maximum of a function is equivariant under translations and invariant under multiplication by a positive number and addition of a constant.

Defining the processes (Z_n) in $C(I_x)$ as

$$\begin{aligned}
Z_n(t) &= 2\delta_n^{-2}\int\left\{\sqrt{z-x}\,1_{[x,\infty)}(z) - \sqrt{z-x-\delta_n t}\,1_{[x+\delta_n t,\infty)}(z)\right\}d(G_n - G)(z)\\
&\quad - \frac{\pi}{4m_0}f(x)t^2,
\end{aligned}$$

we see that

$$\tilde{Z}_n(t) = Z_n(t) + \delta_n^{-2}\left(U(x+\delta_n t) - U(x) - U'(x)\delta_n t - \tfrac{1}{2}U''(x)\delta_n^2 t^2\right).$$

Therefore, using Taylor's theorem, $d(\tilde{Z}_n, Z_n) \to 0$, where d is the metric of uniform convergence on compacta, defined on $C(I_x)$. Therefore, Z_n converges in distribution to some limit process Z in the topology induced by d if and only if $\tilde{Z}_n$ does. We will now establish convergence in distribution of Z_n. It is clear that we only have to concentrate on the part of Z_n containing the randomness:

$$W_n(t) = \delta_n^{-2}\int\left\{\sqrt{z-x}\,1_{[x,\infty)}(z) - \sqrt{z-x-\delta_n t}\,1_{[x+\delta_n t,\infty)}(z)\right\}d(G_n - G)(z),$$

for $t \in I_x$.

For the process W_n the asymptotic convariance structure is given in Lemma 3, the proof of which can be found in the Appendix.

LEMMA 3. *For each fixed* $s, t \in I_x$,

$$\operatorname{Cov}(W_n(s), W_n(t)) = \frac{1}{8}g_0(x)st\left(1 - \frac{\log\log n}{\log n}\right) + O\left(\frac{1}{\log n}\right),$$

provided that $\delta_n = n^{-1/2}\sqrt{\log n}$.

Note that the assumptions made on x imply that $0 < g_0(x) < \infty$. From now on we take $\delta_n = n^{-1/2}\sqrt{\log n}$.

Furthermore, the Lindeberg central limit theorem for triangular arrays yields $W_n(1) \to_{\mathscr{D}} \mathscr{N}(0, g_0(x)/8)$, whereas Chebyshev's inequality implies that, for all fixed $s, t \in I_x$, $|sW_n(t) - tW_n(s)| = o_p(1)$ for $n \to \infty$. Therefore, the finite-dimensional distributions of W_n converge weakly to the corresponding finite-dimensional distributions of the process

$$W(t) = tX \tag{8}$$

in $C(I_x)$, where X is a normally distributed random variable with expectation 0 and variance $\frac{1}{8}g_0(x)$.

Finally, applying the maximal inequality given by Pollard [(1989), Theorem 4.2] to the function class

$$\mathscr{F}_{n,\varepsilon}^{M} = \Big\{\sqrt{z - x - \delta_n t}\, 1_{[x+\delta_n t, \infty)}(z) - \sqrt{z - x - \delta_n s}\, 1_{[x+\delta_n s, \infty)}(z): \\ s, t \in I_x, |s - t| < \varepsilon, \max(|s|, |t|) < M\Big\},$$

with its natural envelope, we obtain the stochastic equicontinuity condition which, together with the convergence in distribution of the finite-dimensional distributions, implies the convergence in distribution in $(C(I_x), d)$ of the sequence (W_n) to W as defined in (8) [Kim and Pollard (1990), Theorem 2.3].

We now use the following lemma, which will be proved in the Appendix.

LEMMA 4. *For $n \to \infty$,*

$$T_n(a_0) = V^{-1}(a_0) + O_p(\delta_n).$$

Note that since the t_n from Kim and Pollard (1990) corresponds to our $\delta_n^{-1}(T_n(a_0) - V^{-1}(a_0))$, Lemma 4 shows that condition (ii) of Theorem 2.7 from Kim and Pollard (1990) is satisfied. Applying Theorem 2.7 from Kim and Pollard (1990) to the process $\tilde{Z}_n$ yields the following result:

$$\begin{aligned}\delta_n^{-1}(T_n(a_0 + \delta_n a) - x) &\to_{\mathscr{D}} \operatorname*{arg\,max}_{t \in I_x} \left\{2tX - \frac{\pi}{4m_0} f(x)t^2 - at\right\} \\ &= \frac{2m_0(2X - a)}{\pi f(x)} 1_{I_x}(2X - a).\end{aligned}$$

Hence,

$$\mathbb{P}\left\{\delta_n^{-1}\left(\hat{V}_n(x) - V(x)\right) \le a\right\} = \mathbb{P}\left\{\delta_n^{-1}(T_n(a_0 + \delta_n a) - x) \le 0\right\} \to \mathbb{P}\{2X \le a\},$$

which completes the proof. □

We get the following as a corollary to Theorem 5.

COROLLARY 2. *Let $x > 0$, and suppose that F has a density f which is strictly positive at x and at 0, and continuous in neighborhoods of x and 0.*

Then

$$\sqrt{\frac{n}{\log n}}\left\{1-\frac{\overline{V}_n(x)}{\hat{V}_n(0)}-F(x)\right\}$$
$$\to_{\mathscr{D}} \mathscr{N}\left(0,\ \frac{g(x)V(0)^2+g(0)V(x)^2}{2V(0)^4}\right)\quad \textit{as } n\to\infty.$$

The proof of this corollary is completely analogous to the proof of Corollary 1, with Theorem 2 replaced by Theorem 5. Note that again, as in Theorem 5, the asymptotic variance is exactly half the asymptotic variance of the naive estimator.

5. Pictures. To give some idea of what the estimators look like, we show the pictures (Figures 1 and 2) of the naive estimator and the isotonized naive estimator of the distribution function F, based on a sample of 100 observations from a standard exponential distribution. In this case, the distribution of the squared circle radii is again standard exponential, so, for simulation purposes, we can just generate a (pseudorandom) sample from a standard exponential distribution and consider this as a sample of squared circle radii, corresponding to a sample of squared sphere radii from a standard exponential distribution.

The pictures (based on the same sample) show that the naive estimator of F roughly has the right trend, apart from the bad behavior between neighboring observation points. The isotonic estimator shows, of course, a much more acceptable behavior.

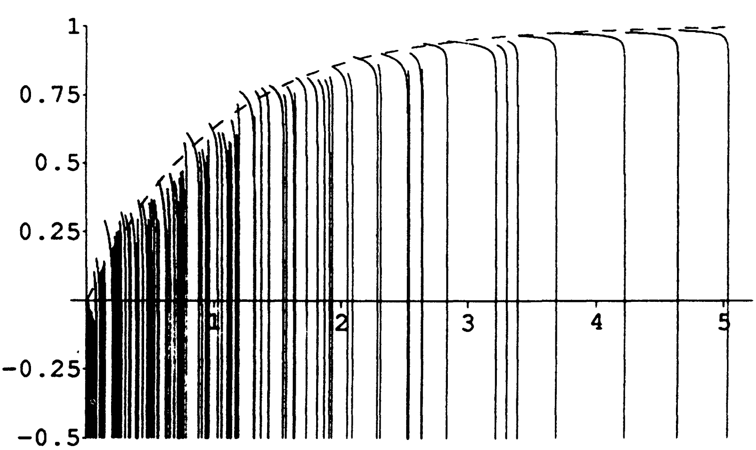

FIG. 1. *The naive estimator of the distribution function of the squared ball radii, based on a sample of size* 100 *from a standard exponential distribution.*

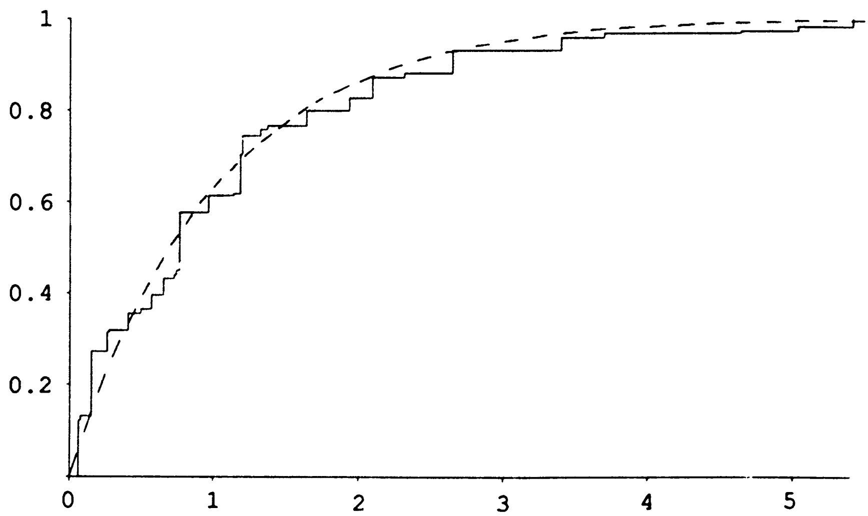

FIG. 2. *The isotonic estimator of the distribution function of the squared ball radii, based on a sample of size* 100 *from a standard exponential distribution.*

APPENDIX

This section consists of the proofs of several lemmas.

PROOF OF LEMMA 1. We define the function $\tilde{V}_\varepsilon$ as follows:

$$\tilde{V}_\varepsilon(x) = V(x) + \int_0^x \phi_\varepsilon^1(y)\,dy.$$

Corresponding to this function, we define the function $\tilde{g}_\varepsilon$ by

$$\tilde{g}_\varepsilon(z) = -\frac{1}{\pi}\int_{(z,\infty)} \frac{d\tilde{V}_\varepsilon(x)}{\sqrt{x-z}}.$$

Note that $\tilde{g}_\varepsilon$ is not a probability density.

By the triangle inequality we have that

$$\|g_\varepsilon - g_0\|_2 \le \|g_\varepsilon - \tilde{g}_\varepsilon\|_2 + \|\tilde{g}_\varepsilon - g_0\|_2. \tag{9}$$

Because the function $(g_\varepsilon - \tilde{g}_\varepsilon)(z)$ has bounded support and tends to zero at rate $\varepsilon/\sqrt{\log 1/\varepsilon}$ uniformly in z, it follows that the first term on the right-hand side of (9) is of order $\varepsilon/\sqrt{\log 1/\varepsilon}$ for $\varepsilon \downarrow 0$. In what follows we will prove that $\|\tilde{g}_\varepsilon - g_0\|_2$ is of order ε.

First of all note that if we consider the $L_2(-\infty,\infty)$-norm of the *extended function* $(\tilde{g}_\varepsilon - g_0)_{\text{ex}}$, that is,

$$(\tilde{g}_\varepsilon - g_0)_{\text{ex}}(z) = -\frac{1}{\pi}\int_{(z,\infty)} \frac{\phi_\varepsilon^1(x)}{\sqrt{x-z}}\,dx \quad \text{for all } z \in \mathbb{R},$$

we have the obvious inequality

$$\|\tilde{g}_\varepsilon - g_0\|_2 \le \|(\tilde{g}_\varepsilon - g_0)_{\mathrm{ex}}\|_2.$$

Notice that $(g_\varepsilon - g_0)_{\mathrm{ex}}$ is the *convolution* of ϕ_ε^1 with the function k defined by

$$k(y) = \frac{1_{(-\infty,0)}(y)}{\sqrt{-y}}.$$

If we denote the Fourier transform of a function f as

$$\mathscr{F}(f)(t) = \hat{f}(t) = \int_{\mathbb{R}} f(z)\exp(itz)\,dz,$$

we can apply Plancherel's identity and use the convolution structure of $(\tilde{g}_\varepsilon - g_0)_{\mathrm{ex}}$ to obtain

$$\|(\tilde{g}_\varepsilon - g_0)_{\mathrm{ex}}\|_2 = \|\mathscr{F}(\tilde{g}_\varepsilon - g_0)_{\mathrm{ex}}\|_2 = \|\hat{k}\cdot\hat{\phi}_\varepsilon^1\|_2.$$

Straightforward calculations yield that

$$\hat{k}(t) = -\sqrt{\frac{\pi}{2}}(1+i)|t|^{-1/2}.$$

For the function $\hat{\phi}_\varepsilon^1$ we have

$$\begin{aligned}\sqrt{\frac{\log 1/\varepsilon}{\varepsilon}}\,\hat{\phi}_\varepsilon^1(t) &= \exp(itx_0)\int_{-\delta}^{\delta}\frac{\sin^2(x/2\varepsilon)}{x}\exp(itx)\,dx\\ &= 2\exp(itx_0)\int_0^{\delta/\varepsilon}\frac{1-\cos x}{x}\sin(t\varepsilon x)\,dx.\end{aligned}$$

Now observe that

$$\int_0^{\delta/\varepsilon}\frac{\sin(tu\varepsilon)}{u}\,du = \int_0^{t\delta}\frac{\sin u}{u}\,du,$$

which is independent of ε and bounded in t. We also have that

$$\begin{aligned}\int_0^{\delta/\varepsilon}\frac{\cos x\sin(t\varepsilon x)}{x}\,dx &= \int_0^{\delta/\varepsilon}\frac{\sin(u+t\varepsilon u)-\sin(u-t\varepsilon u)}{2u}\,du\\ &= \int_{\delta(1/\varepsilon-t)}^{\delta(1/\varepsilon-t)}\frac{\sin u}{2u}\,du.\end{aligned}$$

So, $\sqrt{\log(1/\varepsilon)/\varepsilon}\,\hat{\phi}_\varepsilon^1(t)$ is uniformly bounded both in t and ε.

Now let $\varepsilon \downarrow 0$ and $|t| \to \infty$ in a way that $|t\varepsilon| \to \infty$. Then we have that

$$\begin{aligned}\frac{1}{2}\exp(-itx_0)\sqrt{\frac{\log 1/\varepsilon}{\varepsilon}}\,\hat{\phi}_\varepsilon^1(t) &= \int_0^{\delta/\varepsilon}\frac{1-\cos x}{x}\sin(t\varepsilon x)\,dx\\ &= \int_{\delta t(1-1/t\varepsilon)}^{\delta t}\frac{\sin u}{2u}\,du - \int_{\delta t}^{\delta t(1+1/t\varepsilon)}\frac{\sin u}{2u}\,du.\end{aligned}$$

Because the last two integrals are of order $1/|t\varepsilon|$, we obtain

$$\sqrt{\frac{\log 1/\varepsilon}{\varepsilon}}\,|\hat{\phi}_\varepsilon^1(t)| \sim \frac{1}{|t\varepsilon|} \quad \text{for } |t\varepsilon| \to \infty.$$

Now, write the following for the squared L_2-norm of $\hat{k}\cdot\hat{\phi}_\varepsilon^1$:

$$\int_{\mathbb{R}} \big(\hat{k}(t)\hat{\phi}_\varepsilon^1(t)\big)^2\,dt = \left(\int_{-\infty}^{-1/\varepsilon^2} + \int_{-1/\varepsilon^2}^{1/\varepsilon^2} + \int_{1/\varepsilon^2}^{\infty}\right)\big(\hat{k}(t)\hat{\phi}_\varepsilon^1(t)\big)^2\,dt.$$

Since in the region $|t| > 1/\varepsilon^2$ we have that $|t\varepsilon| \to \infty$ for $\varepsilon \downarrow 0$, we have

$$\left(\int_{-\infty}^{-1/\varepsilon^2} + \int_{1/\varepsilon^2}^{\infty}\right)\big(\hat{k}(t)\hat{\phi}_\varepsilon^1(t)\big)^2\,dt$$

$$\le C\frac{\varepsilon^2}{\log 1/\varepsilon}\int_{1/\varepsilon^2}^{\infty}\frac{1}{t^3\varepsilon^2}\,dt \sim \frac{\varepsilon^4}{\log 1/\varepsilon},$$

for some constant C and for $\varepsilon \downarrow 0$.

Because the function $\sqrt{\log(1/\varepsilon)/\varepsilon}\,\hat{\phi}_\varepsilon^1$ is uniformly bounded and the singularity of $\hat{k}$ at zero is taken care of by the function $\hat{\phi}_\varepsilon^1$ near zero, we may write

$$\int_{-1/\varepsilon^2}^{1/\varepsilon^2}\big(\hat{k}(t)\hat{\phi}_\varepsilon^1(t)\big)^2\,dt \le C\frac{\varepsilon^2}{\log 1/\varepsilon}\int_1^{1/\varepsilon^2}\frac{1}{t}\,dt \sim \varepsilon^2,$$

again for some constant C and for $\varepsilon \downarrow 0$.

Therefore,

$$\|g_\varepsilon - g_0\|_2 = O(\varepsilon) \quad \text{for } \varepsilon \downarrow 0. \qquad \square$$

Proof of Lemma 2. It suffices to show that

$$\int_0^\infty (\phi(x) - \phi^*(x))(\phi^*(x) - \psi(x))\,dx \ge 0,$$

for all $\psi \in \mathscr{F}$. Note that since ϕ is a.e. continuous,

$$\phi(x) - \phi^*(x) = \frac{d(\Phi - \Phi^*)}{dx}(x) \quad \text{a.e.}$$

Therefore, we should prove, for all $\psi \in \mathscr{F}$,

$$\int_0^\infty (\phi^*(x) - \psi(x))\,d(\Phi - \Phi^*)(x) \ge 0.$$

Clearly, by the compact support condition on ϕ, there exists a positive number M such that $\Phi^*(M) = \Phi(M)$ and $\psi^*(x) = \phi(x) = 0$ whenever $x \ge M$. It is also clear that the (negative) measure determined by ϕ^* has all its mass concentrated on the set $\{x > 0: \Phi(x) = \Phi^*(x)\}$. Hence, we may write, using partial integration,

$$\int_0^M (\phi^*(x) - \psi(x))\, d(\Phi - \Phi^*)(x) = (\phi^*(x) - \psi(x))(\Phi(x) - \Phi^*(x))|_0^M$$
$$- \int_0^M (\Phi(x) - \Phi^*(x))\, d(\phi^* - \psi)(x)$$
$$= \int_0^M (\Phi(x) - \Phi^*(x))\, d\psi(x) \geq 0.$$

The partial integration step is justified by the continuity of both Φ and Φ^*. □

Proof of Lemma 3. Fix $s, t \in I_x$. Without loss of generality we may take $s \leq t$.

For $\varepsilon > 0$ we define the function $\phi(\cdot, \varepsilon)$ as follows:

$$\phi(z, \varepsilon) = \sqrt{z - x}\, 1_{[x, \infty)}(z) - \sqrt{z - x - \varepsilon}\, 1_{[x+\varepsilon, \infty)}(z).$$

We can write

$$\operatorname{Cov}(W_n(s), W_n(t)) = \delta_n^{-4} n^{-1} \operatorname{Cov}(\phi(Z, \delta_n t), \phi(Z, \delta_n s)),$$

where Z has density g_0. One easily sees that

$$E\phi(Z, \delta_n t) = O(\delta_n t),$$

implying

$$E\phi(Z, \delta_n s) E\phi(Z, \delta_n t) = O(\delta_n^2 st).$$

We can write

$$\phi(z, \delta_n s)\phi(z, \delta_n t)$$
$$= (z - x)\Bigg(1_{[x, \infty)}(z) + \sqrt{1 - \frac{\delta_n t}{z - x}}\sqrt{1 - \frac{\delta_n s}{z - x}}\, 1_{[x+\delta_n t, \infty)}(z)$$
$$- \sqrt{1 - \frac{\delta_n t}{z - x}}\, 1_{[x+\delta_n t, \varepsilon)}(z) - \sqrt{1 - \frac{\delta_n s}{z - x}}\, 1_{[x+\delta_n s, \infty)}(z)\Bigg).$$

Integrating these four terms using Taylor expansions, gives that the dominant contribution comes from the second term. It follows that

$$E\phi(Z, \delta_n s)\phi(Z, \delta_n t) = -\tfrac{1}{4} st g_0(x)\delta_n^2 \log \delta_n + O(\delta_n^2.)$$

Therefore,

$$\operatorname{Cov}(W_n(s), W_n(t)) = -\tfrac{1}{4} st g_0(x)\delta_n^{-2} n^{-1} \log \delta_n + O(\delta_n^{-2} n^{-1}).$$

Choosing $\delta_n = n^{-1/2}\sqrt{\log n}$, the result follows. □

Proof of Lemma 4. The first thing we will prove is that for all $\varepsilon > 0$ there exists a stochastically bounded sequence of random variables (M_n) and a positive constant R_0 such that

$$
(10) \qquad \begin{aligned} 2\left|\int\left\{\sqrt{z-x}\,1_{[x,\infty)}(z)-\sqrt{z-\theta}\,1_{[\theta,\infty)}(z)\right\}d(G_n-G)(z)\right| \\ \le \varepsilon|x-\theta|^2+\delta_n^2 M_n, \end{aligned}
$$

for all θ with $|\theta - x| < R_0$. Remember that x has been fixed and that $\delta_n = n^{-1/2}\sqrt{\log n}$. This result is similar to Lemma 4.1 in Kim and Pollard (1990) and will be proved along the same lines. Using result (10) we will prove that $T_n(a_0)$ is a δ_n-consistent estimator of $x = V^{-1}(a_0)$. This statement resembles Corollary 4.2 in Kim and Pollard (1990).

For the first part of our proof we introduce the function class

$$
\begin{aligned} \mathscr{H} = \Big\{h_\theta\colon [0,\infty) \to \mathbb{R}\colon h_\theta(z) \\ = 2\sqrt{z-x}\,1_{[x,\infty)}(z) - 2\sqrt{z-\theta}\,1_{[\theta,\infty)}(z) - a_0(\theta - x),\ \theta \ge 0\Big\} \end{aligned}
$$

and, for $R > 0$, its subclass $\mathscr{H}_R$,

$$
\mathscr{H}_R = \{h_\theta \in \mathscr{H}\colon |\theta - x| \le R\}.
$$

It can be proved using arguments similar to those used in the proof of Lemma 3 that there exists a positive number R_0 such that the envelope

$$
H_R(z) = 2\left(\sqrt{z-x-R}\,1_{[x-R,\infty)}(z) - \sqrt{z-x}\,1_{[x,\infty)}(z)\right) + a_0 R
$$

of $\mathscr{H}_R$ satisfies

$$
(11) \qquad \int H_R^2(z) g_0(z)\,dz \le -2g_0(x) R^2 \log R,
$$

for all $R \le R_0$. Now fix $\varepsilon > 0$ and let $(M_n(\omega))$ be the infimum of all positive numbers ν for which

$$
\left|\int h_\theta(z)\,d(G_n - G)(z;\omega)\right| \le \varepsilon|x-\theta|^2 + \delta_n^2 \nu
$$

holds for all $\theta \in [x - R_0, x + R_0]$. If we define, for $n, j \ge 1$,

$$
A(n,j) = \{\theta \ge 0\colon (j-1)\delta_n \le |\theta - x| \le j\delta_n\},
$$

we can write, for all $n \ge 1$ and ν constant,

$$
(12) \qquad \begin{aligned} &\mathbb{P}\{M_n > \nu\} \\ &\le \sum_{j=1}^{R_0/\delta_n} \mathbb{P}\left\{\sup_{\theta\in A(n,j)} \delta_n^{-2}\left|\int h_\theta(z)\,d(G_n-G)(z)\right| > \varepsilon(j-1)^2 + \nu\right\} \\ &\le \sum_{j=1}^{R_0/\delta_n} \frac{\mathbb{E}\left\{\sup_{|\theta - x|\le j\delta_n} \delta_n^{-4}\left[\int h_\theta(z)\,d(G_n - G)(z)\right]^2\right\}}{\left[\varepsilon(j-1)^2+\nu\right]^2}. \end{aligned}
$$

By maximal inequality 3.1.2 in Kim and Pollard (1990) and relation (11), we obtain that there exists a $C > 0$ such that, for each $j \le R_0/\delta_n$,

$$\mathbb{E}\left\{ \sup_{|\theta - x| \le j\delta_n} \delta_n^{-4} \left[\int h_\theta(z)\, d(G_n - G)(z) \right]^2 \right\} \le C g_0(x) j^2.$$

Applying this inequality to (12), it follows that, by choosing ν sufficiently large, $\mathbb{P}\{M_n > \nu\}$ can be make arbitrarily small uniformly in n, which proves (10).

We will now prove that $T_n(a_0)$ is a δ_n-consistent estimator for $x = V^{-1}(a_0)$. Using a Taylor expansion for $\int h_\theta(z)\, dG(z)$ around $\theta = x$, we get that there exists a positive number R_1 such that

$$\int h_\theta(z)\, dG(z) \le -\frac{\pi}{8m_0} f_0(x)(\theta - x)^2,$$

for all θ with $|\theta - x| < R_1$.

Now take ε in relation (10) equal to $\pi f_0(x)/(16m_0)$. Using the fact that $T_n(a_0)$ is weakly consistent for $x = V^{-1}(a_0)$, which follows from Theorem 3, Theorem 4 and relation (6), we know that, with probability tending to 1, $|T_n(a_0) - x| < R_1 \wedge R_0$ for $n \to \infty$. If $|T_n(a_0) - x| < R_1 \wedge R_0$, we obtain that

$$\begin{aligned} 0 = \int h_x(z)\, dG_n(z) &\le \int h_{T_n(a_0)}(z)\, dG_n(z) \\ &\le \int h_{T_n(a_0)}(z)\, dG(z) + \frac{\pi f_0(x)}{16m_0}|x - T_n(a_0)|^2 + \delta_n^2 M_n \\ &\le -\frac{\pi f_0(x)}{16m_0}|x - T_n(a_0)|^2 + \delta_n^2 M_n, \end{aligned}$$

where (M_n) is stochastically bounded. This proves that $T_n(a_0) - x = O_p(\delta_n)$ for $n \to \infty$. □

Acknowledgments. We thank the referees and an Associate Editor for their careful reading of the manuscript and their comments.

REFERENCES

CHOW, Y. S. and TEICHER, H. (1988). *Probability: Independence, Interchangeability and Martingales*. Springer, New York.

GROENEBOOM, P. and WELLNER, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*. Birkhäuser, Boston.

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

JONGBLOED, G. (1991). Nonparametric approach to Wicksell's corpuscle problem. Masters thesis, Faculty of Mathematics and Computer Science, Delft Univ. Technology.

KIM, J. and POLLARD, D. (1990). Cube root asymptotics. *Ann. Statist.* **18** 191–219.

LE CAM, L. M. (1973). Convergence of estimates under dimensionality restrictions. *Ann. Statist.* **1** 38–53.

POLLARD, D. (1989). Asymptotics via empirical processes. *Statist. Sci.* **4** 341–366.

ROBERTSON, T., WRIGHT, F. W. and DYKSTRA, R. L. (1988). *Order Restricted Statistical Inference.* Wiley, New York.

STOYAN, D., KENDALL, W. S. and MECKE, J. (1987). *Stochastic Geometry and Its Applications.* Wiley, New York.

VAN ES, A. J. (1991). *Aspects of Nonparametric Density Estimation. CWI Tract* **77**. Centre for Mathematics and Computer Science, Amsterdam.

WATSON, G. S. (1971). Estimating functionals of particle size distributions. *Biometrika* **58** 483–490.

WICKSELL, S. D. (1925). The corpuscle problem. *Biometrika* **17** 84–99.

WOODROOFE, M. and SUN, J. (1993). A penalized maximum likelihood estimate of $f(0+)$ when f is non-increasing. *Statist. Sinica* **3** 501–515.

DEPARTMENT OF MATHEMATICS
DELFT UNIVERSITY OF TECHNOLOGY
MEKELWEG 4
2628 CD DELFT
THE NETHERLANDS