The Annals of Statistics
1998, Vol. 26, No. 6, 2407–2419

ASYMPTOTICALLY EFFICIENT ESTIMATION IN THE WICKSELL PROBLEM

BY G. K. GOLUBEV AND B. Y. LEVIT

Russian Academy of Sciences and University of Utrecht

We consider the classical Wicksell problem of estimating an unknown distribution function G of the radii of balls, based on their observed cross-sections. It is assumed that the underlying distribution function G belongs to a Hölder class of smoothness $\gamma > 1/2$. We prove that, for a suitable choice of the smoothing parameters, kernel-type estimators are asymptotically efficient for a large class of symmetric bowl-shaped loss functions.

1. Introduction. In the Wicksell problem [Wicksell (1925)] one observes cross-sections, formed by a given plane Π in $\mathbf{R}^3$, of random balls $B_i = B_i(\mathbf{v}_i, R_i)$, $i = 1, \dots, n$, with centers $\mathbf{v}_i$ given by the sites of a stationary Poisson point process, and with iid radii R_i such that $Y_i = R_i^2$ has an unknown distribution function $G(y)$, $y \in \mathbf{R}^+$. We are interested here in estimating $G(y)$ at a given point $y > 0$. For numerous applications of the Wicksell problem in biology, stereology, stochastic geometry and so on, see, for example, Stoyan, Kendall and Mecke (1995). Further references and related problems can be found in Hall and Smith (1988) and the survey paper by Hoogendoorn (1992).

Assume that a coordinate system in $\mathbf{R}^3$ is chosen such that $\Pi = \{\mathbf{v}: v_3 = 0\}$ and denote the observed cross-sections $B_i \cap \Pi$ by $S_i = S(\mathbf{u}_i, r_i)$, $\mathbf{u}_i = (v_{i1}, v_{i2})$. Denote by $F(x)$ the distribution function of the observed squared radii $X_i = r_i^2$ and let λ be the (unknown) intensity of the underlying point process in $\mathbf{R}^3$. Then

$$\begin{aligned}
&\mathbf{P}\{X_1 > x | S_1 \text{ is observed at } v_{11} \in (w_1, w_1 + dw_1), v_{12} \in (w_2, w_2 + dw_2)\}\\
&\quad = \frac{\mathbf{P}\big\{v_{11} \in (w_1, w_1 + dw_1), v_{12} \in (w_2, w_2 + dw_2), |v_{13}| < \sqrt{Y_1 - x}\big\}}{\mathbf{P}\big\{v_{11} \in (w_1, w_1 + dw_1), v_{12} \in (w_2, w_2 + dw_2), |v_{13}| < \sqrt{Y_1}\big\}}\\
&\quad = \frac{2\lambda\, dw_1\, dw_2 \int_x^\infty \sqrt{y - x}\, dG(y)}{2\lambda\, dw_1\, dw_2 \int_0^\infty \sqrt{y}\, dG(y)} = \frac{\int_x^\infty \sqrt{y - x}\, dG(y)}{\int_0^\infty \sqrt{y}\, dG(y)}.
\end{aligned}$$

Hence

$$1 - F(x) = \frac{1}{m(G)} \int_x^\infty \sqrt{y - x}\, dG(y), \tag{1}$$

where

$$m(G) = \int_0^\infty \sqrt{y}\, dG(y)$$

Received October 1996; revised July 1998
AMS 1991 *subject classifications*. Primary 62G05, 62G20; secondary 62C20, 62E20
Key words and phrases. Wicksell's problem, Hölder classes, fractional derivatives, kernel density estimators.

is the expected radius of the balls B_i, which we assume to be finite. Similarly, one can show that $X_1, \dots, X_n$ are independent random variables with distribution function $F(x)$ given by (1). For a more rigorous treatment based on marked point processes see, for example, Reiss (1993), page 47.

One can easily estimate the distribution $F(x)$ of the observed squared radii X_i and then try to "unfold" the unknown distribution $G(y)$ using (1). Problems of this kind are often referred to as inverse problems. The Wicksell problem is obviously also related to the so-called convolution and mixture models.

Fortunately, the underlying equation (1) can in our case be solved explicitly. More precisely, if $G(\cdot)$ is Hölder continuous with an exponent $\gamma > 0$, then

$$1 - G(y) = \frac{2m(G)}{\pi} \int_y^\infty \frac{dF(x)}{\sqrt{x-y}} = \frac{\Phi(y)}{\Phi(0)}, \tag{2}$$

where

$$\Phi(y) = \int_y^\infty \frac{dF(x)}{\sqrt{x-y}}. \tag{3}$$

Although the basic equation (2) has been frequently used in studying the Wicksell problem, the available proofs of it appear to be of a somewhat ad hoc nature. For the reader's convenience we derive (2) here. To find the density of $G(x)$, let $\phi(\cdot)$ be any smooth function having a compact support in $(0, \infty)$. Then according to (1) and changing the order of integration,

$$\begin{aligned}\mathbf{E}\,\phi(X) &= \int_0^\infty \phi'(x)(1 - F(x))\,dx = \frac{1}{m(G)} \int_0^\infty \int_0^y \phi'(x)\sqrt{y-x}\,dx\,dG(y) \\ &= \frac{1}{m(G)} \int_0^\infty \int_0^y \frac{\phi(x)}{2\sqrt{y-x}}\,dx\,dG(y) = \frac{1}{2m(G)} \int_0^\infty \phi(x) \int_x^\infty \frac{dG(y)}{\sqrt{y-x}}\,dx.\end{aligned}$$

Hence the density of F is given by

$$f(x) = \frac{1}{2m(G)} \int_x^\infty \frac{dG(y)}{\sqrt{y-x}}. \tag{4}$$

To express $G(\cdot)$ in terms of $F(\cdot)$, note that according to the above equation,

$$\begin{aligned}\frac{2m(G)}{\pi} \int_y^\infty \frac{dF(x)}{\sqrt{x-y}} &= \frac{1}{\pi} \int_y^\infty \frac{1}{\sqrt{x-y}} \int_x^\infty \frac{dG(z)}{\sqrt{z-x}}\,dx \\ &= \frac{1}{\pi} \int_y^\infty \int_y^z \frac{dx}{\sqrt{(z-x)(x-y)}}\,dG(z) = 1 - G(y)\end{aligned}$$

thus proving (2).

Equations (2) and (3) immediately suggest a naive estimator $\hat{G}_n(y)$ obtained by using the empirical distribution function $\hat{F}_n(x)$ of the observed data $X_1, \dots, X_n$ instead of F. Although such an estimator may have infinite second moments and, being unbounded in y, is very poor in applications, it is asymptotically normal $\mathscr{N}(0, \sigma_0^2(G)/\phi_n^2)$, where $\phi_n^2 = n/\log n$,

$$\sigma_0^2(G) = 4\pi^{-2} m^2(G)\big(f(y) + f(0)(1 - G(y))^2\big)$$

and f is the density of F. Thus the naive estimator is consistent, with a rate of convergence ϕ_n^{-1}. Watson (1971) derives this result as a mere curiosity.

In a recent paper, Groeneboom and Jongbloed (1995) have shown that if $G(\cdot)$ is Hölder continuous with an exponent $\gamma \geq 1$, the above rate of convergence ϕ_n^{-1} cannot be improved by any estimator. They also proposed an isotonic estimator attaining the constant $\sigma_0^2(G)$.

In this paper we derive asymptotically efficient estimators of $G(y)$, under the assumption that $G(\cdot)$ belongs to a given Hölder class Λ_γ, with $\gamma > 1/2$. The efficiency of the proposed estimators holds simultaneously for a large class of symmetric bowl-shaped loss functions, and their mean square error decreases asymptotically as $\sigma^2(G)\phi_n^{-2}$, where $\sigma^2(G) = \sigma_0^2(G)/(2\gamma)$. Such estimators are obtained by replacing the unknown density $f(x)$ of $F(x)$ in (2), (3) by a properly scaled kernel-type density estimator $f_n(x) = f_n(x; X_1, \dots, X_n)$.

The idea of using kernel-type estimators in the Wicksell problem has been proposed by Taylor (1983). Later this method was discussed by Hall and Smith (1988) in the framework of estimating the unknown density $g(\cdot)$ of $G(\cdot)$, by van Es and Hoogendoorn (1990) and by others. The existence of efficient nonparametric estimators in this long-standing problem proved, however, to be elusive for more than 70 years.

To better understand the role played by the Wicksell problem in nonparametric estimation, it is useful to consider a more general family of statistical functionals

$$\Phi_\lambda(y; F) = \int_y^\infty (x - y)^{-\lambda}\, dF(x), \qquad 0 < \lambda < 1, \tag{5}$$

with the special case $\lambda = 1/2$ appearing in (2). Such functionals coincide, up to certain constants, with the Weyl fractional derivatives $D^\lambda F(y)$ of the distribution function F and lead therefore to a natural generalization of the well-known problem of estimating an unknown density and its higher-order derivatives.

For $\lambda < 1/2$, $\Phi_\lambda(y; F)$ is a standard example of a regular statistical functional [see, e.g., Koshevnik and Levit (1976)], since the function $(\cdot - y)^{-\lambda}$ is, normally, square integrable. For $\lambda > 1/2$, on the other hand, $\Phi_\lambda(y; F)$ is a typical example of irregular functionals, exhibiting the same properties as the more familiar density estimation problem. Thus the Wicksell problem (corresponding exactly with the case $\lambda = 1/2$) lies on the *boundary* between regular and irregular nonparametric problems. Moreover it combines, in a peculiar manner, properties of both these types of problems. Namely, asymptotically efficient (locally asymptotically minimax) estimators, although they do exist in the Wicksell problem, are closer in nature to nonparametric density estimators.

2. Main results. Consider a nonparametric estimator of the density $f(\cdot)$ based on the observed random variables $X_1, \dots, X_n$,

$$f_n(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right), \tag{6}$$

where the bandwidth $h = h_n > 0$, $h_n \to 0$ (as $n \to \infty$) and the kernel $K(\cdot)$ will be specified shortly.

Based on $f_n(x)$, a natural estimator of the functional $\Phi(y)$ appearing in (2) is

$$\Phi_n(y) = \int_y^\infty \frac{f_n(x)}{\sqrt{x-y}}\,dx = \int_0^\infty \frac{f_n(y+s)}{\sqrt{s}}\,ds = \frac{1}{n}\sum_{i=1}^n u_h(y - X_i), \tag{7}$$

where

$$u_h(t) = \frac{1}{h}\int_0^\infty K\left(\frac{t+s}{h}\right)\frac{1}{\sqrt{s}}\,ds. \tag{8}$$

Although the distribution of X_i and the target functional $\Phi(y)$ are completely specified by the distribution function F itself, we will endow corresponding expectations, variances and so on with the index G, which is the parameter of primary interest for us [see (1), (2)].

Let

$$G_n(y) = \begin{cases} 0, & \text{if } \Phi_n(y)/\Phi_n(0) > 1, \\ 1 - \Phi_n(y)/\Phi_n(0), & \text{if } \Phi_n(y)/\Phi_n(0) \in [0, 1], \\ 1, & \text{if } \Phi_n(y)/\Phi_n(0) < 0, \end{cases} \tag{9}$$

be the corresponding estimator of $G(y)$ in (2).

Our goal is to show that the thus-defined estimator $G_n(y)$ is asymptotically efficient (or more precisely, locally asymptotically minimax) if the parameter h_n and the kernel $K(\cdot)$ are properly chosen, in accordance with the smoothness properties of the underlying distribution function $G(\cdot)$. To include these properties in our model, consider, for arbitrary $\gamma = \alpha + \beta$, $\alpha = 0, 1, \ldots$, $0 < \beta \le 1$, the Hölder class Λ_γ of functions $g(y)$, $y \ge 0$ having finite norm

$$\|g\|_\gamma = \sup_{y\ge 0} |g(y)| + \sup_{x,\, y \ge 0} \frac{|g^{(\alpha)}(x) - g^{(\alpha)}(y)|}{|x-y|^\beta}.$$

Assuming $\gamma > 1/2$, let $\mathscr{G}_\gamma$ denote the class of all distribution functions $G \in \Lambda_\gamma$ satisfying the following assumptions:

$$G(y) = 0, \qquad y \le 0, \tag{10}$$

$$\int_0^\infty \sqrt{y}\,dG(y) < \infty. \tag{11}$$

Furthermore let $\mathscr{T}_\gamma$ be the topology on $\mathscr{G}_\gamma$ induced by the norm

$$\|G_1 - G_2\|_{\mathscr{G}_\gamma} = \|G_1 - G_2\|_\gamma + \left|\int_0^\infty \sqrt{y}\,d(G_1 - G_2)(y)\right|.$$

In this topological space $(\mathscr{G}_\gamma, \mathscr{T}_\gamma)$ of the underlying distributions, one can consider arbitrary neighborhoods V of a given $G \in \mathscr{G}_\gamma$ as well as limits with respect to a converging net $V \searrow G$.

To incorporate our prior information into the construction of efficient estimators, assume that $K(\cdot)$ is an arbitrary one-sided kernel satisfying the following conditions:

$$\sup_x |K(x)| < \infty, \tag{12}$$

$$\operatorname{supp} K(\cdot) \subset [-1, 0], \tag{13}$$

$$\int K(x)\,dx = 1, \tag{14}$$

$$\int x^i\, K(x)\,dx = 0, \qquad i = 1, \dots, \alpha. \tag{15}$$

Let $\mathscr{W}$ denote the class of loss functions $w(z) \geq 0$, $z \in \mathbf{R}^1$ such that

$$w(z) = w(-z), \qquad w(z) \geq w(y), \qquad |z| \geq |y|,$$

and for some $p, q > 0$ $w(z) \leq p\exp(q|z|)$.

THEOREM 1. *Let $G_n(y)$ be defined by* (6), (7) *and* (9), *where $K(\cdot)$ is an arbitrary kernel satisfying* (12)–(15) *and*

$$h = h_n = n^{-1/(2\gamma)}. \tag{16}$$

If $\gamma > 1/2$ then for any $w \in \mathscr{W}$ and $G_0 \in \mathscr{G}_\gamma$,

$$\lim_{V \to G_0} \limsup_{n \to \infty} \sup_{G \in V} \mathbf{E}_G w\big(\phi_n\big(G_n(y) - G(y)\big)\big) = \mathbf{E}\, w(\xi),$$

where $\phi_n = \sqrt{n/\log n}$ and $\xi \sim \mathscr{N}(0, \sigma^2(G_0))$, with

$$\sigma^2(G) = \frac{2m^2(G)}{\pi^2\gamma}\big(f(y) + f(0)(1 - G(y))^2\big).$$

Next we show that the estimator $G_n(y)$ is locally asymptotically minimax by establishing the following lower bound on the minimax risk.

THEOREM 2. *Let $\gamma > 1/2$. Then for any $w \in \mathscr{W}$ and $G_0 \in \mathscr{G}_\gamma$,*

$$\lim_{V \to G_0} \liminf_{n \to \infty} \inf_{G_n} \sup_{G \in V} \mathbf{E}_G w\big(\phi_n\big(G_n(y) - G(y)\big)\big) \geq \mathbf{E}\, w(\xi),$$

where the infimum is taken over all estimators $G_n(y)$.

The derivation of the lower bound in estimating $G(y)$ requires a slight modification of the well-known technique used in the theory of estimating the so-called regular functionals [see Koshevnik and Levit (1976), Ibragimov and Khasminskii (1981) or a more recent monograph by Bickel, Klaassen, Ritov and Wellner (1993)]. Indeed, as Lemma 1 below shows, the functionals

$\Phi_1(F) = \Phi(y)$ and $\Phi_2(F) = \Phi(0)$ are approximately linear (hence differentiable), with (regularized) gradients $u_h(y - \cdot)$, $u_h(0 - \cdot)$ where $u_h(\cdot)$ and h are defined correspondingly by (8) and (16). Obviously, the gradient of the functional $\Phi(F) = \Phi_1(F)/\Phi_2(F)$ in (2) belongs, for any given F, to the linear span of these two gradients. Let $\mathscr{F}_\gamma$ denote the set of all density functions f corresponding to a distribution F in (1), with some $G \in \mathscr{G}_\gamma$. To obtain the required lower bound it would be sufficient to find, in a vicinity of any given $f_0 \in \mathscr{F}_\gamma$, a two-dimensional family of densities $f_{\mathbf{c}}(x) \in \mathscr{F}_\gamma$, $\mathbf{c} = (c_1, c_2)$, whose score functions form a basis for the above span.

The following parametric family,

$$f_{\mathbf{c}}(x) = f_0(x)\big(1 + c_1\big(u_h(y-x) - \mathbf{E}_{f_0} u_h(y - X_1)\big) \\ + c_2(u_h(-x) - \mathbf{E}_{f_0} u_h(-X_1))\big), \tag{17}$$

would have the desired properties provided that it belonged to the corresponding class of densities $\mathscr{F}_\gamma$.

Unfortunately, there is in general no reason to assume that $f_{\mathbf{c}}(x)$ indeed belongs to this class. However, an approximation of the family (17) introduced below does belong to $\mathscr{F}_\gamma$. Such approximation is based on following formula:

$$\int_u^\infty \frac{1}{t\sqrt{t-u}}\,dt = \pi \mathbf{1}\{u \ge 0\} u^{-1/2}, \tag{18}$$

where for $u < 0$ the integral in (18) is understood in the sense of principal value.

3. Upper bound.

We begin the proof of Theorem 1 with the following preliminary results.

LEMMA 1. *Under the assumptions of Theorem* 1, *for any* $y, z \ge 0$, $y \ne z$ *and* $h \to 0$, *we have*

$$\begin{gathered} \mathbf{E}_G\, u_h(y - X_i) = \Phi(y) + O(h^\gamma), \\ \operatorname{Var}_G u_h(y - X_i) = (f(y) + o(1)) \log(1/h), \\ \operatorname{Cov}_G\big(u_h(y - X_i), u_h(z - X_i)\big) = O(1), \end{gathered} \tag{19}$$

locally uniformly with respect to $G \in \mathscr{G}_\gamma$.

PROOF. Let f denote the density function of the distribution F in (1). Obviously, (2) can be rewritten in terms of $f(x)$ as

$$1 - G(y) = \frac{2m(G)}{\pi} \int_0^\infty \frac{f(y+z)}{\sqrt{z}}\,dz.$$

Using this relation and (14) and (15), one easily obtains

$$
\begin{aligned}
\mathbf{E}_G\, u_h(y - X_i) &= \frac{1}{h}\int f(x)\int_0^\infty K\left(\frac{y-x+s}{h}\right)\frac{1}{\sqrt{s}}\,ds\,dx\\
&= \int K(v)\int_0^\infty f(y-hv+s)\frac{1}{\sqrt{s}}\,ds\,dv\\
&= \frac{\pi}{2m(G)}\int K(v)(1-G(y-hv))\,dv\\
&= \frac{\pi}{2m(G)}(1-G(y)) + \frac{\pi}{2m(G)}\int K(v)(G(y)-G(y-hv))\,dv\\
&= \Phi(y) + O(h^\gamma).
\end{aligned}
$$

Next, it is easy to verify that the density $f(\cdot)$ is equicontinuous on any bounded interval in $\mathbf{R}^+$, locally uniformly in $G \in \mathscr{G}_\gamma$, if $\gamma > 1/2$. The remaining statements of the lemma then follow from the following relations:

$$
\begin{aligned}
&u_h(t) = 0, \qquad t > 0;\\
&\sup_t |u_h(t)| \le Ch^{-1/2};\\
&u_h(t) = (-t)^{-1/2}(1+o(1))
\end{aligned}
\tag{20}
$$

as $h \to 0$, uniformly in $t \le -h(-\log h)^{1/2}$.

The first of these relations is trivial. The second follows from (8) and the inequalities

$$
\begin{aligned}
u_h(t) &= \frac{1}{\sqrt{h}}\int_0^\infty K\left(\frac{t}{h}+s\right)\frac{1}{\sqrt{s}}\,ds\\
&\le \frac{1}{\sqrt{h}}\left(\sup_t |K(t)|\int_0^1 \frac{ds}{\sqrt{s}} + \int_1^\infty K\left(\frac{t}{h}+s\right)ds\right)\\
&\le \frac{1}{\sqrt{h}}\left(2\sup_t |K(t)| + \int |K(t)|\,dt\right).
\end{aligned}
$$

Finally, for the last relation we note that

$$
u_h(t) = \int_{-1}^0 K(s)\frac{1}{\sqrt{hs-t}}\,ds = \int_{-1}^0 K(s)\frac{(1+o(1))}{\sqrt{-t}}\,ds = (1+o(1))(-t)^{-1/2}
$$

uniformly in $t \le -h(-\log h)^{1/2}$ as $h \to 0$.

LEMMA 2. *Under the assumptions of Theorem* 1, *for any* y, $z \ge 0$, $y \ne z$, *locally uniformly with respect to* $G \in \mathscr{G}_\gamma$ *as* $n \to \infty$:

(a) *The random vector* $\phi_n(\Phi_n(y) - \Phi(y), \Phi_n(z) - \Phi(z))$ *is asymptotically normal with zero mean and covariance matrix* $(2\gamma)^{-1}\operatorname{diag}(f(y), f(z))$.

(b) *For any* $\lambda > 0$,

$$
\limsup_{n\to\infty} \mathbf{E}_G \exp\bigl(\lambda\phi_n|\Phi_n(y) - \Phi(y)|\bigr) < \infty. \tag{21}
$$

PROOF. According to (19) and (20), the random variables $u_h(y - X_i)$ in (7) trivially satisfy the Lindeberg condition. Therefore, according to (16) and (19), the random variables $\phi_n(\Phi_n(y) - \Phi(y))$, $\phi_n(\Phi_n(z) - \Phi(z))$ are asymptotically normally distributed $\mathscr{N}(0, f(y)/2\gamma)$, $\mathscr{N}(0, f(z)/2\gamma)$ and asymptotically uncorrelated, locally uniformly with respect to $G \in \mathscr{G}_\gamma$ [cf. Ibragimov and Khasminskii (1981), page 365].

For the second part of Lemma 2, use Bernstein's inequality [see, e.g., Pollard (1984), page 193]; if $Z_1, \dots, Z_n$ are independent identically distributed random variables with $\mathbf{E}Z_i = 0$, $\mathbf{E}Z_i^2 = \sigma^2$ and $|Z_i| \le M$, then for any $x > 0$,

$$\mathbf{P}\left\{\frac{1}{\sigma\sqrt{n}}\left|\sum_{i=1}^{n} Z_i\right| \ge x\right\} \le 2\exp\left[-\frac{x^2}{2}\left(1 + \frac{Mx}{3\sigma\sqrt{n}}\right)^{-1}\right]. \tag{22}$$

It follows from (16), Lemma 1, (20) and (22) that for any $y, \lambda \ge 0$, locally uniformly in $G \in \mathscr{G}_\gamma$,

$$\mathbf{P}_G\{\phi_n|\Phi_n(y) - \Phi(y)| > x\} = O(e^{-\lambda x}) \tag{23}$$

for $x, n \to \infty$, thus proving (21). □

LEMMA 3. *For any $y, \lambda \ge 0$, locally uniformly with respect to $G \in \mathscr{G}_\gamma$ as $n \to \infty$,*

$$\limsup_{n\to\infty} \mathbf{E}_G \exp(\lambda\phi_n|G_n(y) - G(y)|) < \infty.$$

PROOF. Let $A = A_n = \{X_1, \dots, X_n\colon |\Phi_n(0) - \Phi(0)| \le \Phi(0)/2\}$. Then, according to (2), (16), Lemma 1 and (20)–(22), for some $C > 0$ we have

$$\begin{aligned}
&\mathbf{E}_G \exp(\lambda\phi_n|G_n(y) - G(y)|)\\
&\quad = \mathbf{E}_G \exp(\lambda\phi_n|G_n(y) - G(y)|)(\mathbf{1}_A + \mathbf{1}_{A^c})\\
&\quad \le \mathbf{E}_G \exp(C\lambda\phi_n(|\Phi_n(y) - \Phi(y)| + |\Phi_n(0) - \Phi(0)|))\mathbf{1}_A\\
&\qquad + \exp(\lambda\sqrt{n})\mathbf{E}_G \mathbf{1}_{A^c}\\
&\quad \le \mathbf{E}_G \exp(2C\lambda\phi_n|\Phi_n(y) - \Phi(y)|) + \mathbf{E}_G \exp(2C\lambda\phi_n(|\Phi_n(0) - \Phi(0)|))\\
&\qquad + \exp(\lambda\sqrt{n} - Cn\sqrt{h}) = O(1).
\end{aligned}$$
□

PROOF OF THEOREM 1. Due to (2) and Lemma 2,

$$\begin{aligned}
G_n(y) - G(y) &= -\frac{\Phi_n(y) - \Phi(y)}{\Phi(0)} + \frac{\Phi(y)}{\Phi^2(0)}(\Phi_n(0) - \Phi(0)) + o_{\mathbf{P}}(\phi_n^{-1})\\
&= \frac{2m(G)}{\pi}\Big(-(\Phi_n(y) - \Phi(y)) + (1 - G(y))(\Phi_n(0) - \Phi(0))\Big)\\
&\quad + o_{\mathbf{P}}(\phi_n^{-1}).
\end{aligned}$$

Therefore, according to Lemma 2, the random variable $\phi_n(G_n(y) - G(y))$ is asymptotically normal $\mathscr{N}(0, \sigma^2(G))$, locally uniformly in $G \in \mathscr{G}_\gamma$.

This together with Lemma 3 implies Theorem 1 in case of a continuous loss function $w \in \mathscr{W}$. To complete the proof, note that any function $w \in \mathscr{W}$ can be approximated by a sequence of continuous functions $w_\delta \in \mathscr{W}$, $\delta \to 0$, such that $w(|x|) \le w_\delta(|x|) \le w(|x| + \delta)$ [e.g., by

$$w_\delta(x) = \frac{1}{\delta} \int K\left(\frac{|x| - y}{\delta}\right) w(y)\, dy,$$

where $K(\cdot) \ge 0$ is a continuous function satisfying (13) and (14))]. □

4. The lower bound. According to the discussion at the end of Section 2, in order to obtain the required lower bound one has to exhibit the existence of an (asymptotically least favorable) parametric subfamily $G_{\mathbf{c}}$, in a vicinity of a given distribution $G_0 \in \mathscr{G}_\gamma$, such that the corresponding family of densities $f_{\mathbf{c}}$ due to (4) approximates (17). Such a least favorable parametric subfamily passing through a given $G_0 \in \mathscr{G}_\gamma$ is proposed below. However, for the construction to work, the initial distribution G_0 should itself be sufficiently *rough*, that is, satisfy the following conditions:

$$\liminf_{x \searrow 0} G(x)/x^{\gamma \wedge 1} > 0, \tag{24}$$

$$\liminf_{x \to y} |G(x) - G(y)|/|x - y|^{\gamma \wedge 1} > 0. \tag{25}$$

Evidently, not every distribution $G_0 \in \mathscr{G}_\gamma$ satisfies these assumptions. However, in any vicinity V of a given G_0 such distribution functions do exist. It is assumed throughout this section that $y > 0$ is fixed.

Let $\varepsilon > 0$. For $\gamma \ge 1$ choose any distribution function $G_0^\varepsilon \in \mathscr{G}_\gamma$ satisfying (24)–(25) such that

$$\left\| G_0 - G_0^\varepsilon \right\|_{\mathscr{G}_\gamma} \le \varepsilon. \tag{26}$$

For $1/2 < \gamma < 1$ and $\varepsilon < y$, let

$$G_0^\varepsilon(u) = \left(1 + 3\gamma^{-1}(1-\gamma)\varepsilon\right)^{-1}\left(G_0(u) + \int_0^u (g^\varepsilon(v) + g^\varepsilon(v - y))\, dv\right), \tag{27}$$

where

$$g^\varepsilon(x) = \max\left(|x/\varepsilon|^{\gamma - 1} - 1, 0\right). \tag{28}$$

Here again G_0^ε belongs to $\mathscr{G}_\gamma$ and satisfies (24) and (25).

Further, let $K(\cdot)$ be a kernel satisfying (13) and (14) and such that $\|K^{(-1)}\|_\gamma < \infty$, where $K^{(-1)}(x) = \int_{-\infty}^x K(s)\, ds$. For any sufficiently small $h > 0$ and $\delta = 1/\sqrt{\log(1/h)}$, let

$$\chi(t) = t^{-1} \mathbf{1}\{h \le |t| \le \delta\}, \tag{29}$$

$$\bar{\chi}(t) = \frac{1}{h} \int K\left(\frac{t - s}{h}\right) \chi(s)\, ds. \tag{30}$$

Below we consider the following parametric subfamily of distribution functions:

$$G_{\mathbf{c}}(u) = \frac{1}{D_{\mathbf{c}}}\Big(G_0^{\varepsilon}(u) + \int_0^u \big(c_1\bar{\chi}(v) + c_2\bar{\chi}(v-y)\big)\,dv\Big), \tag{31}$$

where

$$D_{\mathbf{c}} = 1 + \int_0^{\infty} \big(c_1\bar{\chi}(v) + c_2\bar{\chi}(v-y)\big)\,dv.$$

LEMMA 4. *Let $G_0 \in \mathscr{G}_{\gamma}$. Then for any vicinity $V \subset \mathscr{G}_{\gamma}$ of G_0 there exists $\varepsilon > 0$ such that for all $0 < h < \varepsilon$ and $|c_i| < \varepsilon h^{\gamma}$:*

(a) $G_{\mathbf{c}} \in V$;
(b) *As* $h \to 0$,

$$G_{\mathbf{c}}(y) = G_0^{\varepsilon}(y) - c_1\big(1 - G_0^{\varepsilon}(y)\big)\log h + c_2 \log h + o\big(\|\mathbf{c}\| \log h\big)$$

and

$$m(G_{\mathbf{c}}) = \frac{m(G_0^{\varepsilon})}{1 + c_1 \log h}\big(1 + O(\|\mathbf{c}\|)\big). \tag{32}$$

PROOF. It can be easily seen that the distribution function G_0^{ε} belongs to V for all sufficiently small ε. According to (29) and (30), $\sup_t |\bar{\chi}(t)| = O(1/h)$ and, for all $u, v \geq 0$,

$$\begin{aligned} \big|\bar{\chi}^{(\alpha-1)}(u) - \bar{\chi}^{(\alpha-1)}(v)\big| &\leq h^{-\alpha}\int_0^{\infty}\left|K^{(\alpha-1)}\Big(\frac{u}{h} - t\Big) - K^{(\alpha-1)}\Big(\frac{v}{h} - t\Big)\right| dt \\ &\leq Ch^{-\gamma}|u - v|^{\beta}. \end{aligned} \tag{33}$$

Statement (a) follows now for $1/2 < \gamma < 1$ from (27), (31), (33) and the fact that $g^{\varepsilon}(t) > c\bar{\chi}(t)$ for all $|c| \leq \varepsilon h^{\gamma}$. For $\gamma > 1$ one can similarly use (33) and the assumptions (24) and (25).

Part (b) of Lemma 4 follows from Taylor's formula and the following easily verifiable relations [cf. (29), (30)]:

$$\int_0^y \bar{\chi}(t-y)\,dt = (1 + o(1))\log h, \qquad \int_0^{\infty} \bar{\chi}(t)\,dt = -(1 + o(1))\log h,$$
$$\int_0^{\infty} \bar{\chi}(t-y)\,dt = O(1). \qquad \square$$

Our next task is to describe the behavior of the family of distributions $F_{\mathbf{c}}(x)$ associated with the family $G_{\mathbf{c}}(x)$ in (31) according to (1). In particular we shall substantiate the claim that the corresponding densities $f_{\mathbf{c}}(x)$ provide a good approximation to (17). Consider the functions

$$\zeta(x) = \int_x^{\infty} \frac{\chi(y)\,dy}{\sqrt{y-x}}, \qquad \bar{\zeta}(x) = \int_x^{\infty} \frac{\bar{\chi}(y)\,dy}{\sqrt{y-x}} \tag{34}$$

appearing in $f_{\mathbf{c}}(x)$; compare (36). The properties of $\bar{\zeta}(x)$ are described in the following lemma.

LEMMA 5.

$$\bar{\zeta}(x)=\begin{cases}0, & \text{if } x>\delta,\\ \pi x^{-1/2}+O(\delta^{-1/2}), & \text{if } h\le x\le\delta,\\ O(h^{-1/2}), & \text{if } -h\le x\le h,\\ O(h(-x)^{-3/2})+O(\delta^{-1/2}), & \text{if } -\delta\le x\le -h,\\ O(\delta^{-1/2}), & \text{if } x\le-\delta.\end{cases}$$

To prove Lemma 5 one should first show that it holds with $\bar{\zeta}$ replaced by ζ. [Note that $\zeta(x)$ in (34) can be obtained in a closed form; cf. (18)]. The result for $\bar{\zeta}$ then follows by elementary calculus.

PROOF OF THEOREM 2. Recall that a family of measures $\mathbf{P}_{\mathbf{c}}^{(n)}$, $\mathbf{c}\in\Theta_n\subset\mathbf{R}^s$, is called *locally asymptotically normal* at the point $\mathbf{c}=0$ with a normalizing factor $\psi_n\to\infty$, if:

(i) for any $\mathbf{v}\in\mathbf{R}^s$, $\mathbf{v}/\psi_n\in\Theta_n$ for all sufficiently large n;

(ii) there exists a sequence of functions $\Delta_n=\Delta_n(X^{(n)})$, $X^{(n)}=(X_1,\dots,X_n)$ and a positive definite matrix I, such that for any $\mathbf{v}\in\mathbf{R}^s$,

$$\log\frac{d\mathbf{P}_{\mathbf{v}/\psi_n}^{(n)^c}}{d\mathbf{P}_0^{(n)}}(X^{(n)})=\mathbf{v}^\top\Delta_n-\frac{\mathbf{v}^\top I\mathbf{v}}{2}+o_{\mathbf{P}}(1),$$

where $\Delta_n\sim_{\mathbf{P}_0^{(n)}}\mathscr{N}(0,I)$ [cf. Ibragimov and Khasminskii (1981), Section II.2].

A useful application of local asymptotic normality is provided by the following result [see, e.g., Ibragimov and Khasminskii (1981), Section II)]. Let a family $\mathbf{P}_{\mathbf{c}}^{(n)}$ satisfy condition (ii) and:

(i′) For any compact $K\subset\mathbf{R}^s$, $K/\psi_n\subset\Theta_n$ for all sufficiently large n.

Then for any nonnegative symmetric quasi-convex loss function $W(u)$, $u\in\mathbf{R}^s$,

$$\liminf_{n\to\infty}\inf_{T_n}\sup_{\mathbf{c}\in\Theta_n}\mathbf{E}_{\mathbf{c}}W\big(\psi_n(T_n-\mathbf{c})\big)\ge\mathbf{E}\,W(\xi), \tag{35}$$

where $T_n=T_n(X^{(n)})$ is an arbitrary estimator of $\mathbf{c}$ and $\xi\sim\mathscr{N}(0,I^{-1})$.

Let $\Theta_n=\{\mathbf{c}\colon |c_1|,|c_2|\le\varepsilon h^\gamma\}$. Choose $h=h_n$ according to (16) in the definition of the family $G_{\mathbf{c}}(\cdot)$ [see (29)–(31)] and let $\psi_n=\sqrt{n\log n}$. Clearly, condition (i′) above is satisfied with this choice of h; see Lemma 4. Using a Taylor expansion in (32) gives

$$\begin{aligned} f_{\mathbf{c}}(x)&=\frac{1}{2m(G_{\mathbf{c}})}\int_x^\infty\frac{dG_{\mathbf{c}}(y)}{\sqrt{y-x}}\\ &=\big(1+O(\|\mathbf{c}\|)\big)f_0^\varepsilon(x)+c_1(1+o(1))\frac{\bar{\zeta}(x)}{2m(G_0^\varepsilon)}+c_2(1+o(1))\frac{\bar{\zeta}(x-y)}{2m(G_0^\varepsilon)},\end{aligned} \tag{36}$$

uniformly in x, where

$$f_0^\varepsilon(x) = \frac{1}{m(G_0^\varepsilon)} \int_x^\infty \frac{dG_0^\varepsilon(y)}{\sqrt{y-x}}.$$

Further, the well-known argument used in proving local asymptotic normality for independent observations (sustained by Lemma 5) and a bound on the remainder terms using, for example, the Bernstein inequality (22) show that the family $\mathbf{P}_{\mathbf{c}}^{(n)}$ corresponding to densities $f_{\mathbf{c}}(x)$ in (36) is locally asymptotically normal, with the matrix I given by

$$\frac{\pi^2}{8\gamma m^2(G_0^\varepsilon)} \operatorname{diag}\left(\frac{1}{f_0^\varepsilon(0)}, \frac{1}{f_0^\varepsilon(y)}\right).$$

Therefore, according to Lemma 4(b), for any uniformly continuous loss function $w \in \mathscr{W}$, we have

$$\begin{aligned} r(V) &= \liminf_{n\to\infty} \inf_{G_n} \sup_{G\in V} \mathbf{E}_G\, w\bigl(\phi_n\bigl(G_n(y) - G(y)\bigr)\bigr) \\ &\geq \liminf_{n\to\infty} \inf_{G_n} \sup_{G_{\mathbf{c}}} \mathbf{E}_{G_{\mathbf{c}}}\, w\bigl(\phi_n(G_n(y) - G_{\mathbf{c}}(y))\bigr) \\ &\geq \liminf_{n\to\infty} \inf_{T_n} \sup_{\mathbf{c}} \mathbf{E}_{\mathbf{c}}\, w\bigl((2\gamma)^{-1}\psi_n\bigl(T_n - c_1(1 - G_0^\varepsilon(y)) + c_2\bigr)\bigr). \end{aligned}$$

An application of (35) to the loss function

$$W(x_1, x_2) = w\bigl((2\gamma)^{-1}(x_1(1 - G_0^\varepsilon(y)) - x_2)\bigr)$$

results in the lower bound $r(V) \geq \mathbf{E}\, w(\xi)$, where $\xi \sim \mathscr{N}(0, \sigma^2(G_0^\varepsilon))$. Note that according to (26) and (27), $\lim_{\varepsilon\to 0} \sigma^2(G_0^\varepsilon) = \sigma^2(G_0)$, thus proving Theorem 2 in the case of a uniformly continuous loss function w. Finally, the proof of Theorem 2 is completed by approximating an arbitrary loss function $w \in \mathscr{W}$ by a sequence of uniformly continuous functions $w_\delta \in \mathscr{W}$ with $w_\delta \nearrow w$. $\square$

Acknowledgment. We thank an anonymous referee for numerous comments on an earlier version of the paper, which led to an improved exposition of the material.

REFERENCES

BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1993). *Efficient and Adaptive Inference in Semiparametric Models*. Johns Hopkins Univ. Press.

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

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

HOOGENDOORN, A. W. (1992). Estimating the weight undersized distribution for the Wicksell problem. *Statist. Neerlandica* **4** 259–282.

IBRAGIMOV, I. A. and KHASMINSKII, R. Z. (1981). *Statistical Estimation. Asymptotic Theory.* Springer, New York.

KOSHEVNIK, YU. A. and LEVIT B. Y. (1976). On a non-parametric analog of the information matrix. *Theory Probab. Appl.* **21** 738–753.

POLLARD, D. (1984). *Convergence of Stochastic Processes.* Springer, New York.

REISS, R.-D. (1993). *A Course on Point Processes*. Springer, Berlin.

STOYAN, D., KENDALL, W. S. and MECKE, J. (1995). *Stochastic Geometry and Its Applications*, 2nd ed. Wiley, New York.

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

VAN ES, B. and HOOGENDOORN, A. W. (1990). A kernel approach to estimation of the sphere radius density in Wicksell's corpuscle problem. *Biometrika* **77** 139–145.

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

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

INSTITUTE OF PROBLEMS
OF INFORMATION TRANSMISSION
RUSSIAN ACADEMY OF SCIENCE
BOLSHOI KARETNYI 19
101447 MOSCOW
RUSSIA

INSTITUTE OF MATHEMATICS
UNIVERSITY OF UTRECHT
BUDAPESTLAAN 6
3584 CD UTRECHT
THE NETHERLANDS