Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

A.J. van Es, A.W. Hoogendoorn

A kernel approach to estimation of the sphere radius density in Wicksell's Corpuscle Problem

Department of Mathematical Statistics Report MS-R8809 June

The Centre for Mathematics and Computer Science is a research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a nonprofit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

A KERNEL APPROACH TO ESTIMATION OF THE SPHERE RADIUS DENSITY IN WICKSELL'S CORPUSCLE PROBLEM

A.J. van Es
Mathematisch Instituut
Universiteit van Amsterdam
Roetersstraat 15
1018 WB Amsterdam

A.W. Hoogendoorn
Centrum voor Wiskunde en Informatica
Kruislaan 413
1098 SJ Amsterdam

We consider the estimation of the probability density function of the radii of spheres in a medium given the radii of their profiles in a random slice. This problem is known as Wicksell's corpuscle problem. We present an estimator related to the classical kernel density estimator and discuss its properties. A comparison is made with two other estimators related to the kernel estimator.

AMS 1980 subject classification: 62G05
Key words and Phrases: cross-validation, kernel density estimators, particle size distribution, stereology

1. Introduction.

Suppose that we want to estimate the probability density f of the radius of a random sphere in an opaque medium such as for instance a drop of oil in a piece of rock, and suppose that we have a sample of spheres which we can not observe directly. Instead we do have a sample $X_1,\ldots, X_n$ of radii of n circular profiles obtained by taking slices from the spheres. The estimation of the sphere radius density f from the sample of profile radii is known as the *corpuscle problem.* This problem was first considered by Wicksell (1925) who, under suitable assumptions on the way the slices are obtained, derived the following two relations between the density g of the radii of the circular profiles $X_1,\ldots, X_n$ and the density f of the radii of the spheres. If we assume that both the supports of f and g are contained in the interval [0,R] for some positive constant R then we have

$$g(y) \quad = \quad \frac{1}{\mu} \int_y^R \frac{y}{\sqrt{r^2 - y^2}} f(r)dr, \qquad 0 < y \leq R, \tag{1.1}$$

Report MS-R8809
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

and

$$1 - F(r) = \frac{2\mu}{\pi} \int_r^{\infty} \frac{1}{\sqrt{y^2 - r^2}} g(y)dy, \quad r \geq 0, \tag{1.2}$$

where F denotes the distribution function of the sphere radii and μ is equal to its expected value, i.e.

$$\mu = \int_0^R rf(r)dr. \tag{1.3}$$

Taking r equal to zero in relation (1.2) we see that we also have

$$\mu = \frac{\pi}{2}\left(\int_0^{\infty} \frac{1}{y} g(y)dy\right)^{-1}. \tag{1.4}$$

Above we assume that we are dealing with a *sample* of circle radii, in particular that the observations are independent. In many practical cases this will not be true but it can be shown that relations (1.1) and (1.2) still hold under more realistic assumptions. Throughout this paper however we shall not consider such situations and we assume that the observations are independent. For reviews of the Wicksell problem see for instance Ripley (1981) and Stoyan, Kendall & Mecke (1987). In recent stereological litterature it is emphasized that having data on close, parallel sections allows much more simple and much more nicely behaved estimators. However in practical work this is still not always possible.

Concerning estimation of the density f of the sphere radii several parametric and nonparametric methods have been proposed. In this paper we confine ourselves to nonparametric methods and to methods related to the classical Parzen-Rosenblatt kernel estimator in particular. To indicate the variety of other ways to deal with the Wicksell Problem we mention the following methods. Anderssen & Jakeman (1975) derive an estimator for f by numerically solving the equation obtained by differentiating (1.2). They use Lagrange smoothing of the empirical distribution function (on a chosen grid), product integration and spectral differentiation. In its simplest form (the trapezoidal case, using the datapoints as gridpoints) this boils down to substitution of the cumulative frequency polygon. Nychka, Wahba, Goldfarb & Pugh (1984) use a penalized least squares method (or cross-validated spline method) to find an estimator for f. Their solution is a cubic spline on a chosen grid. Wilson (1987) uses a smoothed EM-algorithm (EMS), considering the Wicksell problem as a problem with incomplete data. Her method combines the EM-algorithm with simple smoothing because just applying the EM-algorithm gives very unsmooth solutions.

In section 2 we introduce an estimator of the density of the sphere radii which is related to the kernel estimator. This estimator can be obtained from an estimator of the distribution function of the sphere radii studied by Watson (1971). In section 4 we suggest a cross-validation method which can possibly be used to determine suitable bandwidths for this estimator. Two other estimators related to the kernel estimator, proposed in Taylor (1983) and Hall & Smith (1988), are given in section 3. In the remainder of the paper these three estimators are compared by a simulation study and theoretical results concerning their expectations and variances. For the properties of kernel estimators we refer to the recent monographs of Devroye & Györfi (1985), Silverman (1986) and Devroye (1987).

2. An estimator of the density of the sphere radii.

Before we present our estimator of the density of the sphere radii we first recall some properties of the kernel estimator. If $X_1,\dots, X_n$ is a sample from a distribution with a density f then the kernel estimator of this density is given by

$$f_{nh}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right), \tag{2.1}$$

where K is a probability density function called the *kernel* and h is a positive smoothness parameter called the *window* or the *bandwidth*. One of the ways the kernel estimator can be obtained is to take the derivative of a smoothed empirical distribution function F_n. A simple calculation shows

$$f_{nh}(x) = \frac{d}{dx}\int_{-\infty}^{\infty} \frac{1}{h} K\left(\frac{x-s}{h}\right) F_n(s)ds. \tag{2.2}$$

We shall need this formula later but first we proceed with some other properties of the kernel estimator. It follows that the expectation of $f_{nh}(x)$ is independent of the sample size n, we have

$$E\, f_{nh}(x) = \int_{-\infty}^{\infty} \frac{1}{h} K\left(\frac{x-s}{h}\right) f(s)ds. \tag{2.3}$$

If f is twice differentiable and if K is symmetric then it can be shown under some additional regularity conditions that we have for h tending to zero

$$E\, f_{nh}(x) = \int_{-\infty}^{\infty} K(s)f(x+hs)ds =$$

$$\int_{-\infty}^{\infty} K(s)(f(x) + hsf'(x) + \tfrac{1}{2}h^2s^2f''(x) + \dots)ds =$$

$$f(x) + \tfrac{1}{2}h^2f''(x)\int_{-\infty}^{\infty} s^2K(s)ds + \dots .$$

Notice that we have explicitly used the fact that K is symmetric, otherwise the second term would have been of order h and thus the bias would have been larger. On the other hand a standard computation shows that the variance of $f_{nh}(x)$ is of order 1/nh. The usual conclusions drawn from these observations is that we should use sequences of bandwidths (h_n) satisfying $h_n \to 0$ and $nh_n \to \infty$, and that bandwidths which for smooth densities asymptotically minimize the mean integrated squared error

$$\mathrm{MISE}_n(h) := \int_{-\infty}^{\infty} (f_{nh}(x) - f(x))^2 dx =$$

$$\int_{-\infty}^{\infty} (Ef_{nh}(x) - f(x))^2 dx + \int_{-\infty}^{\infty} \mathrm{var}(f_{nh}(x)) dx \tag{2.4}$$

are of order $n^{-1/5}$.The best possible mean integrated squared error is of order $n^{-4/5}$. This is seen from the fact that the integrated squared variance is of order 1/nh and from the fact that the integrated squared bias, i.e. the first term in (2.4), for smooth densities f is of order h^4.

It follows from the previous remarks that if we want to construct a kernel type estimator of the sphere radius density f we would like the expectation to be equal to (2.3). Taking the kernel K equal to a differentiable probability density function and defining the estimator $f_{nh}(x)$ for x>0 by

$$f_{nh}(x) := \frac{-2}{\pi n h^2} \sum_{i=1}^{n} \int_0^{X_i} \frac{1}{\sqrt{X_i^2 - u^2}} K'\left(\frac{x-u}{h}\right) du, \tag{2.5}$$

the next theorem shows that the expectation of this estimator, apart from an unknown factor $1/\mu$ and a correction term, equals the expectation of a kernel estimator *based on n observations of sphere radii.* This means that we can use f_{nh} as an estimator of the function f/μ.

Theorem 2.1. *If* K *is a differentiable probability density function with a bounded derivative and if* f *is a bounded density with a support contained in* $[\varepsilon, R]$ *for some* $0<\varepsilon<R<\infty$ *then we have for* h>0

$$E\, f_{nh}(x) = \frac{1}{\mu} \int_{-\infty}^{\infty} \frac{1}{h} K\left(\frac{x-s}{h}\right) f(s) ds - \frac{1}{\mu h} K\left(\frac{x}{h}\right), \tag{2.6}$$

where μ *is given by* (1.3).

Proof. Notice that the conditions imposed on f and relation (1.1) imply that g is a bounded density with a support contained in [0,R]. By Fubini's theorem we can rewrite the expectation of $f_{nh}(x)$ as follows,

$$E\, f_{nh}(x) =$$

$$E \frac{-2}{\pi h^2} \int_0^{X_i} \frac{1}{\sqrt{X_i^2 - u^2}} K'\left(\frac{x-u}{h}\right) du =$$

$$\frac{2}{\pi} \int_0^{\infty} \left(\int_0^{t} \frac{1}{\sqrt{t^2 - u^2}} \frac{-1}{h^2} K'\left(\frac{x-u}{h}\right) du \right) g(t) dt =$$

$$\int_0^{\infty} \left(\frac{2}{\pi} \int_u^{\infty} \frac{1}{\sqrt{t^2 - u^2}} g(t) dt \right) \frac{-1}{h^2} K'\left(\frac{x-u}{h}\right) du.$$

Next by formula (1.2) we see that this integral equals

$$\frac{1}{\mu} \int_0^{\infty} (1 - F(u)) \frac{-1}{h^2} K'\left(\frac{x-u}{h}\right) du =$$

$$\frac{1}{\mu} \int_0^{\infty} \left(\int_u^{\infty} f(s) ds \right) \frac{-1}{h^2} K'\left(\frac{x-u}{h}\right) du =$$

$$\frac{1}{\mu} \int_0^{\infty} \left(\int_0^{s} \frac{-1}{h^2} K'\left(\frac{x-u}{h}\right) du \right) f(s) ds =$$

$$\frac{1}{\mu} \int_0^{\infty} \left(\frac{1}{h} K\left(\frac{x-s}{h}\right) - \frac{1}{h} K\left(\frac{x}{h}\right) \right) f(s) ds =$$

$$\frac{1}{\mu} \int_0^{\infty} \frac{1}{h} K\left(\frac{x-s}{h}\right) f(s) ds - \frac{1}{\mu h} K\left(\frac{x}{h}\right),$$

which completes the proof of the theorem. □

Concerning the correction term in (2.6) notice that if xK(x) converges to zero for x tending to infinity we have

$$\frac{1}{\mu h}K(\frac{x}{h}) \to 0, \text{ as } h\downarrow 0, \tag{2.7}$$

for every positive x. If K is a density with support [-1,1], which we shall assume from now on, this term is equal to zero for all $x \geq h$.

The next result gives an upperbound for the variance of the estimator (2.5). Notice that this bound is larger than the variance of the usual kernel estimator which is of order 1/nh. The proof of this theorem is left to the appendix.

Theorem 2.2. *If* K *is a symmetric differentiable probability density function with a bounded derivative and with support* [-1,1], *and if* f *satisfies the conditions of theorem 2.1 then we have*

$$\mathrm{var}(f_{nh}(x)) = \frac{g(x)}{x}O\left(\frac{-\log(h)}{nh^2}\right) + \frac{1}{x}o\left(\frac{-\log(h)}{nh^2}\right), \text{ as } h\downarrow 0, \tag{2.8}$$

for all $x>0$, *where the order bounds are independent of* x.

Just as in (2.2) the estimator (2.5) can be derived by taking the derivative of a smoothed estimator of the distribution function F of the sphere radii. Let G_n denote the empirical distribution function of the sample of circle radii $X_1,\ldots,X_n$. From formula (1.2) we then obtain an estimator F_n of F by plugging in G_n. We get

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n}\left(\left(1 - \frac{2\mu(X_1,\ldots,X)}{\pi}\frac{1}{\sqrt{X_i^2 - x^2}}\right)I_{[0,X_i)}(x) + I_{[X_i,\infty)}(x)\right) =$$
$$I_{[0,\infty)}(x) - \frac{2\mu(X_1,\ldots,X_n)}{\pi n}\sum_{i=1}^{n}\frac{1}{\sqrt{X_i^2 - x^2}}I_{[0,X_i)}(x),$$

where

$$\mu(X_1,\ldots,X_n) = \pi n \Big/ \sum_{i=1}^{n}\frac{2}{X_i}. \tag{2.9}$$

This is an unsatisfactory estimator of F since it is not monotonous and has values out of [0,1]. However, Watson (1971) showed that in spite of these bad properties F_n is a consistent estimator of F. We obtain our estimator from F_n as follows,

$$\frac{d}{dx}\int_{-\infty}^{\infty}\frac{1}{h}K\left(\frac{x-s}{h}\right)F_n(s)ds =$$

$$\frac{d}{dx}\int_{0}^{\infty}\frac{1}{h}K\left(\frac{x-s}{h}\right)ds - \frac{2\mu(X_1,\ldots,X_n)}{\pi n}\sum_{i=1}^{n}\frac{d}{dx}\int_{-\infty}^{\infty}\frac{1}{h}K\left(\frac{x-s}{h}\right)\frac{1}{\sqrt{X_i^2 - s^2}}I_{[0,X_i)}(s)ds =$$

$$\frac{1}{h}K\left(\frac{x}{h}\right) - \frac{2\mu(X_1,\dots,X_n)}{\pi n h^2}\sum_{i=1}^{n}\int_0^{X_i}\frac{1}{\sqrt{X_i^2 - s^2}}K'\left(\frac{x-s}{h}\right)ds$$

which is equal to our estimator f_{nh} except for the estimate $\mu(X_1,\dots,X_n)$ of μ and the first term.

Remark 2.3. A direct consequence of theorems 2.1 and 2.2 is that in order to get a consistent estimator of $f(x)/\mu$ for a fixed $x>0$ we have to use sequences of bandwidths (h_n) satisfying $h_n \to 0$ and $-nh_n^2/\log(h_n) \to \infty$. There is no need to worry about the factor $g(x)/x$ in (2.8) since under our conditions on f it can be shown that $g(x)/x$ converges to a finite constant as x tends to zero. The variance is discussed further in section 6.

Remark 2.4. It is readily checked that (2.6) also holds if K is a continuous density function which has a bounded derivative except in finitely many points. Thus we are allowed to use the Epanechnikov kernel

$$K(x) = \frac{3}{4}(1-x^2)I_{[-1,1]}(x), \tag{2.10}$$

which has a well known optimality property with respect to the mean integrated squared error criterion. However, it should be noted that, since the variance of our estimator is not the same as the variance of the usual kernel estimator, this optimality property does not hold here. An advantage of this kernel is that we have explicit expressions for the integrals in (2.5) needed to compute the estimator so we don't need numerical integration.

Remark 2.5. The usual argument against higher order kernels, i.e. the fact that the corresponding kernel estimator can become negative at some points does not hold here since even if we use positive kernels our estimator can also have a negative value at some points.

3. Two other estimators related to the kernel estimator.

Above we have seen that our estimator (2.5) can be obtained by first transforming the empirical distribution function G_n of the circle radii into an estimator of the distribution function F of the circle radii, which is then followed by smoothing and differentiation to obtain a kernel type estimator. For two previous estimators this transforming and smoothing is interchanged. Taylor (1983) first estimates the density g of the circle radii by a kernel estimator and then transforms the estimate by formula (1.2) into an estimate of the density f. If g_{nh} is a kernel estimator of the density g then the estimator is given by

$$f_{nh}^T(x) = -\frac{2\mu}{\pi}\frac{d}{dx}\int_x^{\infty}\frac{1}{\sqrt{y^2 - x^2}}g_{nh}(y)dy. \tag{3.1}$$

Hall & Smith (1988) base their estimator on the squares of the radii. Let f_1 and g_1 denote the densities of the squares of the sphere radii and the circle radii then we have $f_1(x)=(2x^{1/2})^{-1}f(x^{1/2})$ and $g_1(x)=(2x^{1/2})^{-1}g(x^{1/2})$. It turns out that the relation between these densities has a convolution structure which makes them easier to work with. The estimator $f_{nh}^{(1)}$ of f_1 is a transformed kernel estimate of g_1 based on the sample $X_1^2,\ldots,X_n^2$, i.e.

$$f_{nh}^{(1)}(x) = -\frac{2\mu}{\pi}\frac{d}{dx}\int_x^\infty \frac{1}{\sqrt{y-x}} g_{nh}^{(1)}(y)dy, \tag{3.2}$$

where $g_{nh}^{(1)}(y)$ is given by

$$g_{nh}^{(1)}(y) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{y-X_i^2}{h}\right). \tag{3.3}$$

An estimator of f is then obtained by

$$f_{nh}^{HS}(x) = 2xf_{nh}^{(1)}(x^2). \tag{3.4}$$

Above we have assumed that the value of μ is known. Since μ depends on the unknown density we either have to be satisfied with an estimator of f/μ, which we get by omitting the factor μ in (3.1) and (3.2), or we have to estimate μ. A straightforward estimator of μ is given by (2.9), but since $1/X_i$ does not have a finite variance it does not converge at a rate of $\sqrt{n}$. However, Hall and Smith show that the rate of convergence does come arbitrarily close to $\sqrt{n}$, which is faster than the rate of convergence of the density estimators given above. Therefore this estimator can be used without disturbing the asymptotics. The same estimator of μ can be used to construct an estimator of f from our estimator (2.5).

4. Selecting the bandwidth by least squares cross-validation.

Let us again assume that the support of f is contained in an interval $[\varepsilon,R]$, $0<\varepsilon<R<\infty$, and let f_{nh} denote the estimator (2.5). A major problem in kernel estimation is the choice of the smoothing parameter h. Suppose that we would like to use a bandwidth h which minimizes the mean integrated squared error

$$\mathrm{MISE}_n(h) := E\int_\varepsilon^\infty \left(f_{nh}(x) - \frac{f(x)}{\mu}\right)^2 dx,\ h>0, \tag{4.1}$$

or at least one which asymptotically minimizes this error, then this bandwidth depends on the unknown density and is therefore unknown itself. In this section we suggest a least squares cross-validation method to compute a suitable bandwidth from the data. The method is similar to the least squares cross-validation bandwidth selection method for the ordinary kernel estimator.

We proceed as follows. Notice that minimizing (4.1) is equivalent to minimizing

$$\mathrm{MISE}_n(h) - \frac{1}{\mu^2}\int_\varepsilon^\infty f^2(x)dx = E\int_\varepsilon^\infty f_{nh}^2(x)\,dx - \frac{2}{\mu}\int_\varepsilon^\infty E\, f_{nh}(x)\, f(x)\, dx. \tag{4.2}$$

In order to construct an estimator of (4.2) we define the *leave one out estimator* $f_{nh}^{(j)}(x)$ by

$$f_{nh}^{(j)}(x) := \frac{-2}{\pi(n-1)h^2}\sum_{i=1,i\neq j}^{n}\int_0^{X_i}\frac{1}{\sqrt{X_i^2-u^2}}K'\Big(\frac{x-u}{h}\Big)du,\ j = 1,\dots,n, \tag{4.3}$$

i.e. the estimator (2.5) based on the observations $X_1,\dots, X_n$ with X_j left out. Next we define the random function L_n

$$L_n(h) := \int_\varepsilon^\infty f_{nh}^2(x)\,dx - \frac{4}{\pi n}\sum_{j=1}^{n} I_{[\varepsilon,\infty)}(X_j)\int_\varepsilon^{X_j}\frac{1}{\sqrt{X_j^2-x^2}}\frac{d}{dx}f_{nh}^{(j)}(x)\,dx,\ h>0. \tag{4.4}$$

The following result shows that the expectation of $L_n(h)$ equals (4.2) apart from a term $2Ef_{nh}(\varepsilon)/\mu$.

Theorem 4.1. *If* K *is a symmetric probability density function with a bounded second derivative and support* [-1,1] *and if* f *satisfies the conditions of theorem 2.1 then we have*

$$E\,L_n(h) = E\int_\varepsilon^\infty f_{nh}^2(x)\,dx - \frac{2}{\mu}\int_\varepsilon^\infty E\, f_{nh}(x)\, f(x)\,dx + \frac{2}{\mu}\,Ef_{nh}(\varepsilon). \tag{4.5}$$

Proof. We compute the expectation of the second term in (4.4) as follows. By $Ef_{nh}^{(j)}(x) = Ef_{nh}(x)$ and the transformation formula (1.2) we have

$$E\frac{4}{\pi n}\sum_{j=1}^{n} I_{[\varepsilon,\infty)}(X_j)\int_\varepsilon^{X_j}\frac{1}{\sqrt{X_j^2-x^2}}\frac{d}{dx}f_{nh}^{(j)}(x)\,dx =$$

$$\frac{4}{\pi}\int_\varepsilon^\infty\int_\varepsilon^{t}\frac{g(t)}{\sqrt{t^2-x^2}}\frac{d}{dx}E\,f_{nh}^{(j)}(x)\,dxdt =$$

$$\frac{4}{\pi}\int_\varepsilon^\infty\Big(\int_x^\infty\frac{g(t)}{\sqrt{t^2-x^2}}dt\Big)\frac{d}{dx}E\,f_{nh}(x)\,dx =$$

$$\frac{2}{\mu}\int_\varepsilon^\infty (1 - F(x)) \frac{d}{dx} E f_{nh}(x)\, dx =$$

$$-\frac{2}{\mu}(1 - F(\varepsilon))\, Ef_{nh}(\varepsilon) + \frac{2}{\mu}\int_\varepsilon^\infty E\, f_{nh}(x)\, f(x)\, dx =$$

$$-\frac{2}{\mu}\, Ef_{nh}(\varepsilon) + \frac{2}{\mu}\int_\varepsilon^\infty E\, f_{nh}(x)\, f(x)\, dx,$$

which proves (4.5). □

The results above suggest that we might obtain a good bandwidth if we minimize the criterion function $LS_n(h)$ defined by

$$LS_n(h) := L_n(h) - \frac{2f_{nh}(\varepsilon)}{\hat{\mu}_n}, \tag{4.6}$$

where $\hat{\mu}_n$ is suitable estimator of μ such as for instance (2.9). A straightforward computation shows that this function can also be written as

$$LS_n(h) = \int_\varepsilon^\infty f_{nh}^2(x)\, dx - \frac{2f_{nh}(\varepsilon)}{\hat{\mu}_n} + \frac{8}{\pi^2(n-1)nh^3} \sum_{i\neq j} I_{[\varepsilon,\infty)}(X_j) \int_0^{X_i}\int_\varepsilon^{X_j} \frac{1}{\sqrt{X_i^2 - x^2}} \frac{1}{\sqrt{X_j^2 - u^2}} K''\left(\frac{x-u}{h}\right) dx du. \tag{4.7}$$

Remark 4.2. If h is smaller than ε, which implies that the correction term (2.7) in (2.6), is equal to zero on $[\varepsilon,\infty)$, and if f is equal to zero on the interval $[0,\varepsilon+h)$ then it is readily seen that $E\, f_{nh}(\varepsilon)$ vanishes. This implies that in that case the term $2f_{nh}(\varepsilon)/\hat{\mu}_n$ in (4.6) can be omitted.

Since we have not proved that this method works, which would require a more detailed analysis than just a computation of the expectation of L_n, we should be very careful when we actually use it to compute a bandwidth. This should be stressed even more because the least squares cross-validation method for the ordinary kernel estimator is known to have slow convergence properties (see Hall & Marron (1987a, 1987b)). We have not performed simulations with this bandwidth selection method since we could not find a kernel such that numerical integration in the computation of the last term of (4.7) can be avoided. Without such a kernel evaluation of the cross-validation criterion seems to require a lot of computing time.

5. Simulations.

To avoid the problem of estimating μ we consider the three methods as methods to estimate f/μ instead of f. As before let f_{nh} denote the estimator (2.5), let f_{nh}^T denote the estimator given by (3.1) with the factor μ omitted, and let f_{nh}^{HS} denote the estimator of f/μ obtained by omitting the factor μ in (3.2). The main objective of the simulation study is to compare these three estimators.

The sphere radius density f in our simulations is equal to a mixture of two normal densities conditioned to be positive. The density of the mixture is equal to $0.7\,\phi_1(x) + 0.3\,\phi_2(x)$, where ϕ_1 is a normal density with mean 0.15 and standard deviation 0.03 and ϕ_2 is a normal density with the same standard deviation but with mean 0.275. The same density, but truncated on the left at 0.04, together with two other densities, was used by Wilson (1987) in her simulation study. For the generation of the samples of circle radii we have to be aware of the fact that their density is equal to $f_c(x)=xf(x)/\mu$ instead of f. This is caused by the fact that spheres with a large radius have a higher probability to be cut by the slice, and so they have a higher probability to appear in the sample of profile radii. To generate samples from the density f we have used random number generators from the IMSL package. Next, to obtain samples of observations with density f_c, we have used a rejection technique , see for instance Ripley (1987) p.60. The kernel used in the computation of the estimators is the Epanechnikov kernel (2.10). For this kernel the integrals in the definition of the estimators can be derived analytically.

Figures 1.a, 1.b, and 1.c contain the graphs of the estimators f_{nh}, f_{nh}^T and f_{nh}^{HS}, computed from the same sample of size 500. The graph of f/μ is denoted by the dotted line. In figure 1.a the continuous line is the estimator f_{nh} plus an estimate of the correction factor in (2.6), i.e.

$$f_{nh}(x) + \frac{1}{\mu(X_1,\ldots,X_n)} K\left(\frac{x}{h}\right), \tag{5.1}$$

where $\mu(X_1,\ldots,X_n)$ is given by (2.9). For $x \geq h$ the corrected and uncorrected estimates are equal, since then the second term vanishes. The dashed line on $[0,h]$ denotes the uncorrected estimate f_{nh} on this interval. The same is repeated in figures 2, 3 and 4 for different samples of size 500, and in figures 5, 6, 7 and 8 for different samples of size 1000. The bandwidths were chosen by eye.

The estimates of μ for these samples are given in the next table. The true value of μ is 0.1875.

Fig.	n	estimate of μ
1	500	0.1809
2	500	0.1812
3	500	0.1868
4	500	0.1966
5	1000	0.1874
6	1000	0.1820
7	1000	0.1865
8	1000	0.1973

Table 1. Estimates of μ for the samples of figures 1, ..., 8.

The mean of the four estimates computed from the samples of size 500 is 0.1864 and the standard deviation is 0.0073. For the four estimates computed from the samples of size 1000 the mean is 0.1883 and the standard deviation is 0.0064.

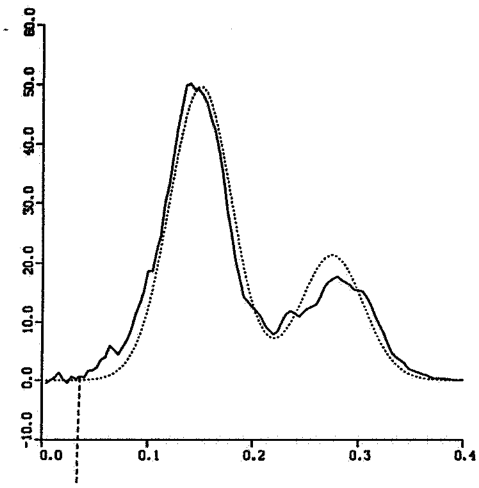

Figure 1.a. f_{nh}, n=500, h=0.036.

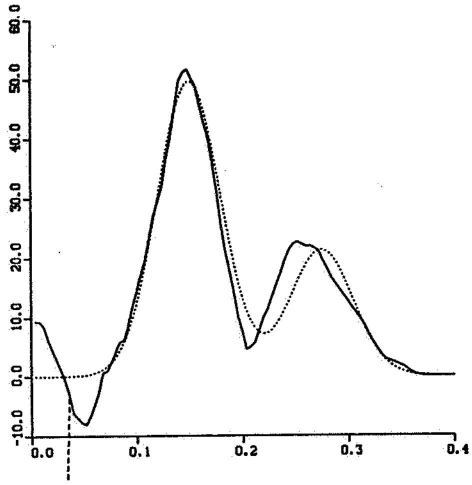

Figure 2.a. f_{nh}, n=500, h=0.036.

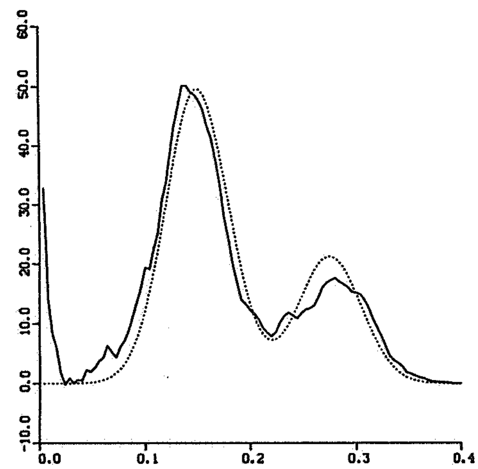

Figure 1.b. f^{T}_{nh}, n=500, h=0.036.

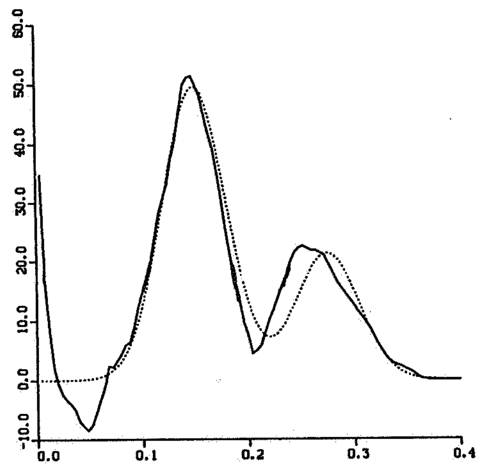

Figure 2.b. f^{T}_{nh}, n=500, h=0.036.

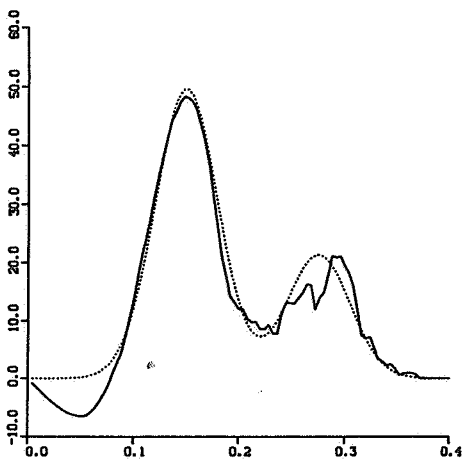

Figure 1.c. f^{HS}_{nh}, n=500, h=0.012.

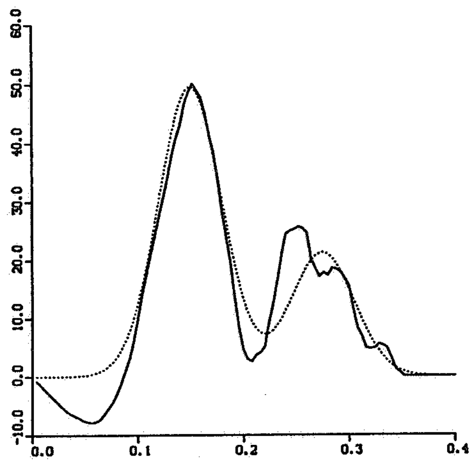

Figure 2.c. f^{HS}_{nh}, n=500, h=0.012.

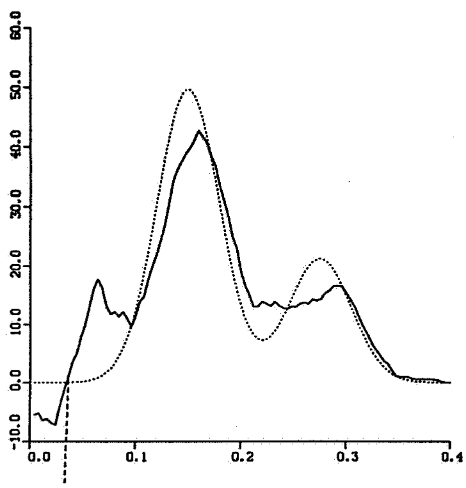

Figure 3.a. f_{nh}, n=500, h=0.036.

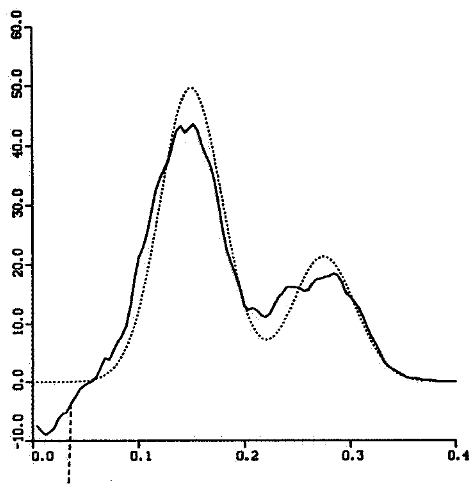

Figure 4.a. f_{nh}, n=500, h=0.036.

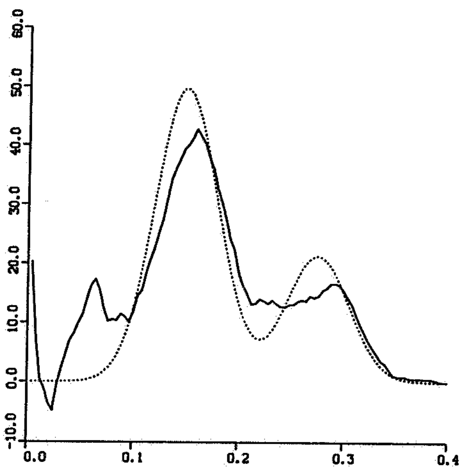

Figure 3.b. f^T_{nh}, n=500, h=0.036.

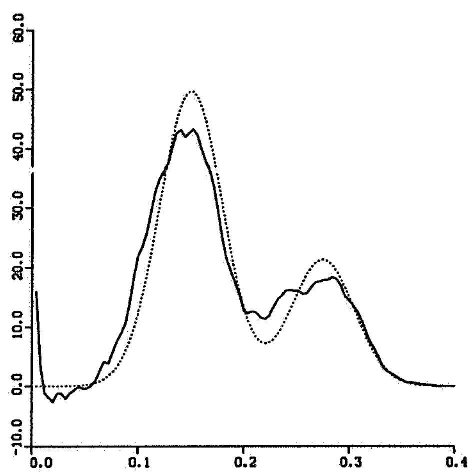

Figure 4.b. f^T_{nh}, n=500, h=0.036.

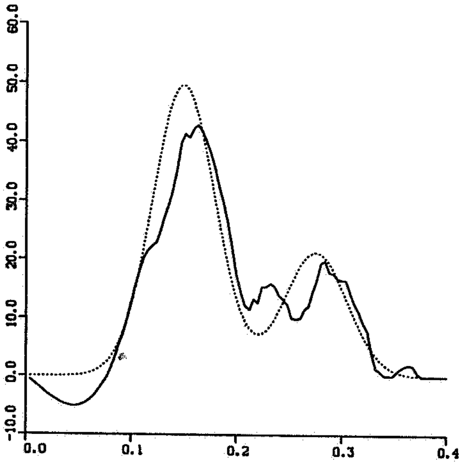

Figure 3.c. f^{HS}_{nh}, n=500, h=0.012.

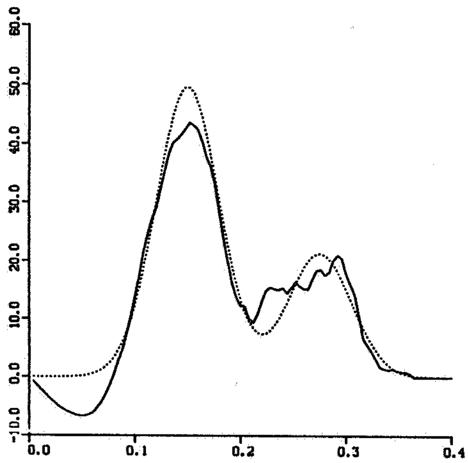

Figure 4.c. f^{HS}_{nh}, n=500, h=0.012.

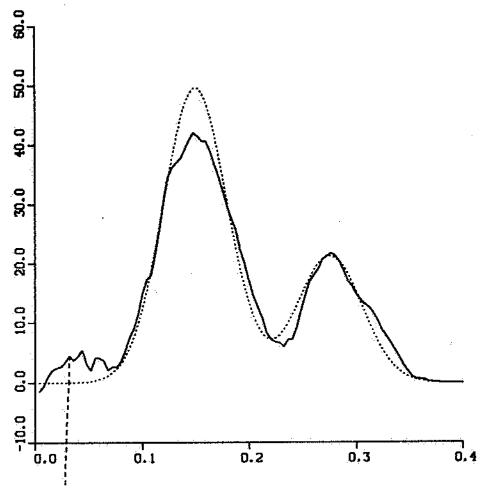

Figure 5.a. f_{nh}, n=1000, h=0.03.

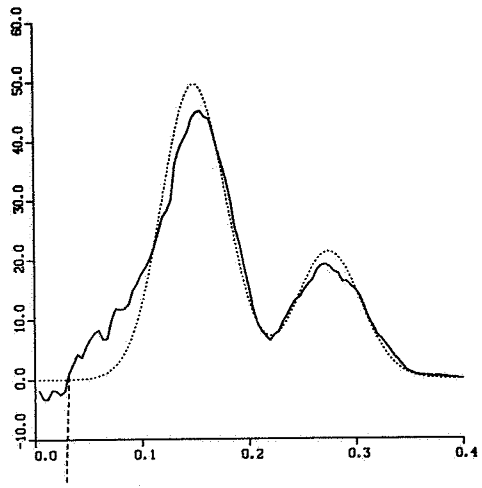

Figure 6.a. f_{nh}, n=1000, h=0.03.

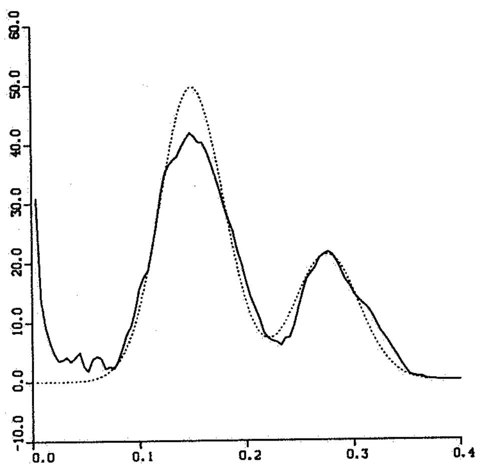

Figure 5.b. f^{T}_{nh}, n=1000, h=0.03.

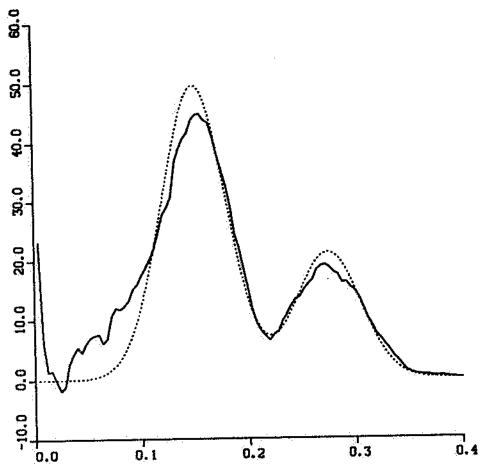

Figure 6.b. f^{T}_{nh}, n=1000, h=0.03.

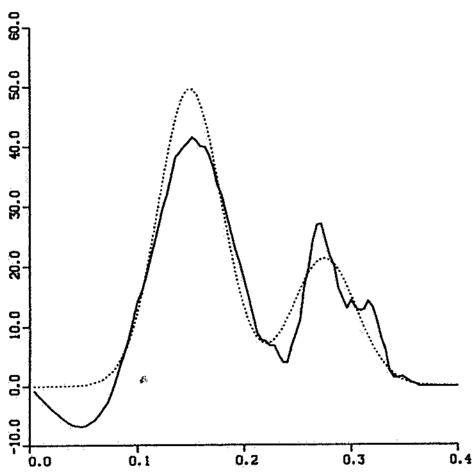

Figure 5.c. f^{HS}_{nh}, n=1000, h=0.01.

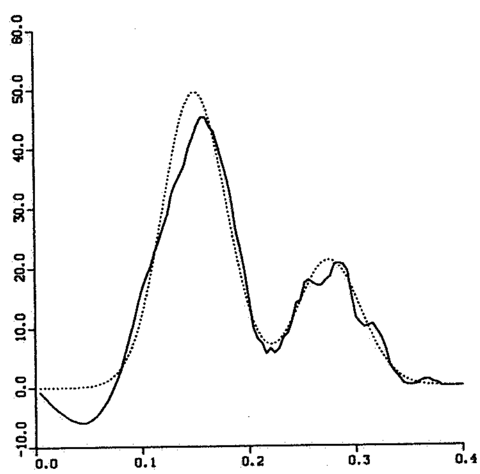

Figure 6.c. f^{HS}_{nh}, n=1000, h=0.01.

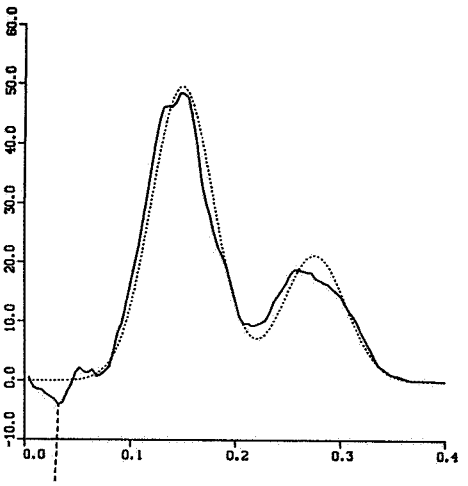

Figure 7.a. f_{nh}, n=1000, h=0.03.

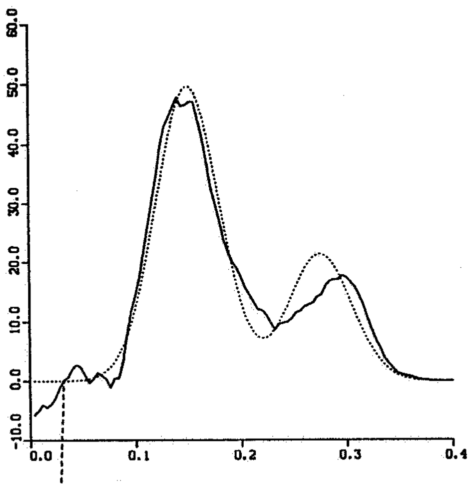

Figure 8.a. f_{nh}, n=1000, h=0.03.

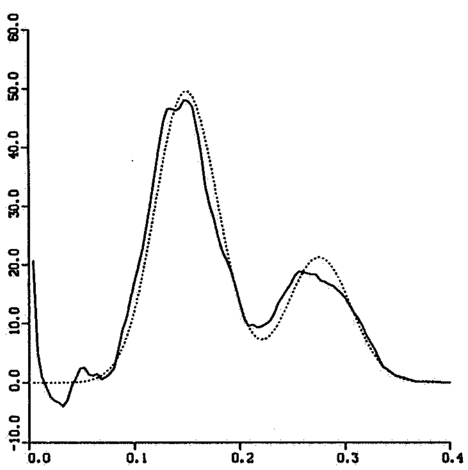

Figure 7.b. f^{T}_{nh}, n=1000, h=0.03.

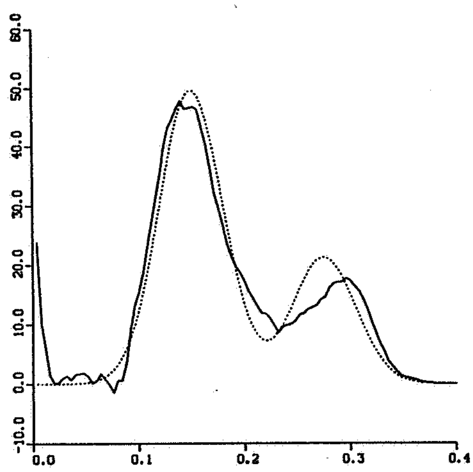

Figure 8.b. f^{T}_{nh}, n=1000, h=0.03.

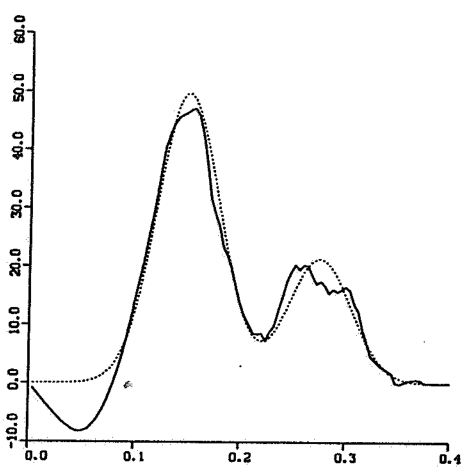

Figure 7.c. f^{HS}_{nh}, n=1000, h=0.01.

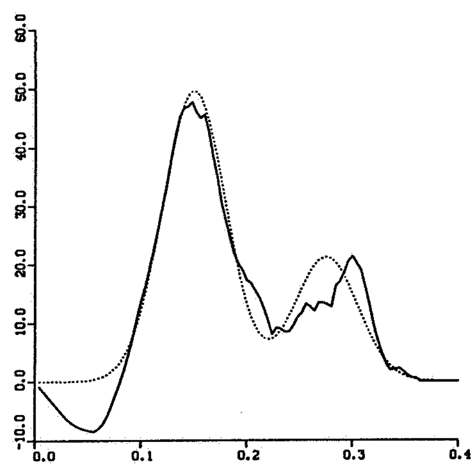

Figure 8.c. f^{HS}_{nh}, n=1000, h=0.01.

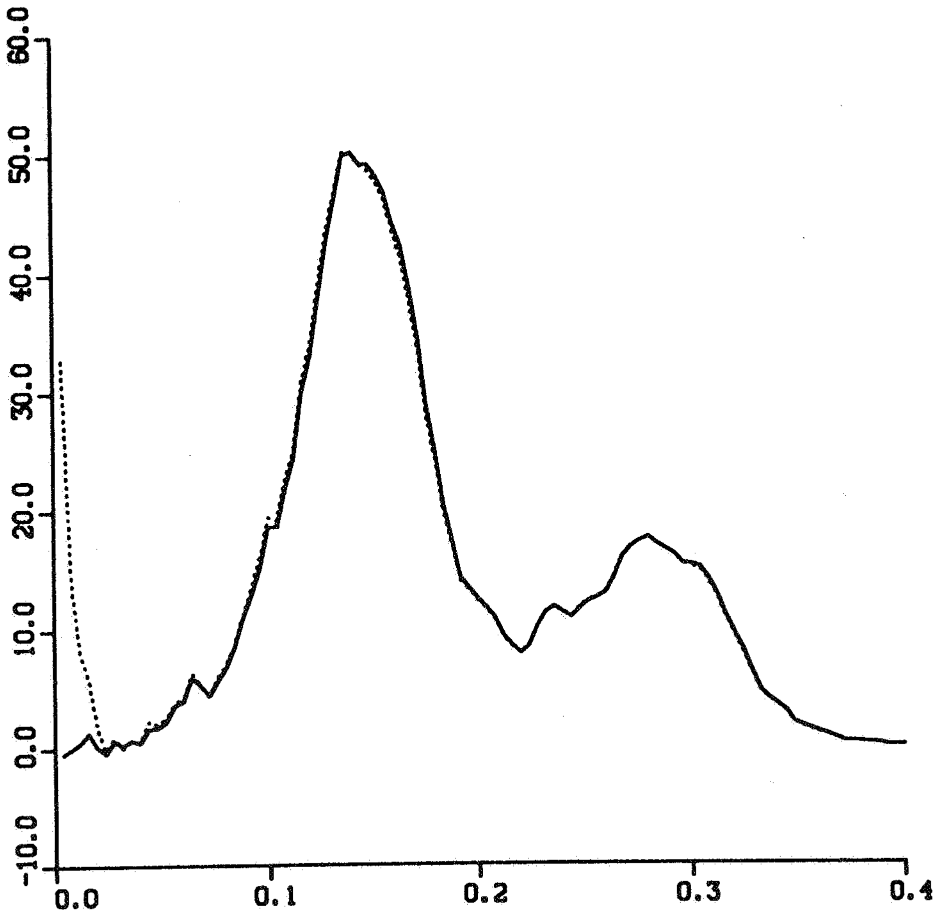

Figure 9. The estimates f_{nh} and f^T_{nh} of figures 1.a and 1.b. n=500. h=0.036.

In figure 9 we have drawn f_{nh} and f^T_{nh} for the sample of figures 1.a and 1.b. The picture shows that there is practically no difference between these two estimators except close to zero.

6. Discussion.

The most striking things about the estimator f_{nh} given by (2.5) are the appearance of the correction term in its expectation (2.6) and its remarkably short derivation. In the asymptotic expansions of the expectation of Taylor's estimator and their own estimator in Hall & Smith (1984, 1988) such a term is absent. However, the simulation results in the previous section indicate that these estimators also have systematic large errors close to zero. This can be explained from the fact that, however smooth the density f may be, the density g of the circle profile radii will always have a kink in zero. Moreover, the density g_1 of the squared circle radii will have a jump in zero, which is clear from the relation $g_1(x)=(2x^{1/2})^{-1}g(x^{1/2})$. If we use kernels with support [-1,1] then this sort of discontinuities in f and its derivative in zero will cause a kernel estimator to have a large bias on the interval [-h,h]. On that interval the bias will be of larger order than the order h^2, which is the order of the bias for smoother densities. For Taylor's estimator f_{nh}^{T} this means that the estimator of the circle radius density will have a larger bias on [0,h], which in its turn implies that the bias of f_{nh}^{T} will also be large on this interval. Similarly a kernel estimator of g_1 will also have a large bias on [0,h], so f_{nh}^{HS} will have a large bias on $[0,h^{1/2}]$. The impact of discontinuities of f and its derivative on the bias of kernel estimators is treated in detail in Van Es (1988).

Another important observation is that, except for small samples and large bandwidths, the estimators f_{nh} and f_{nh}^{T} are practically equal on the interval $[h,\infty)$. Since Hall & Smith (1984,1988) show that the asymptotic variance of Taylor's estimator is of order $1/nh^2$, this also means that the upper bound on the variance of f_{nh} given by (2.8) is probably not sharp, and that the factor log(h) can be omitted. However, we have not been able to derive the exact asymptotic variance.

Supposing that the asymptotic variance of our estimator is indeed of order $1/nh^2$, just as the variances of f_{nh}^{T} and f_{nh}^{HS}, which is also shown in Hall & Smith (1988), this means that the asymptotically optimal bandwidths for the mean integrated squared error for these estimators are of order $n^{-1/6}$, and that the resulting best possible error is of order $n^{-2/3}$. Here we should be a little bit more careful and define the mean integrated squared error as the expectation of the squared error, integrated over an interval $[c,\infty)$, where c is some positive constant, otherwise the large bias close to zero causes a larger error. It is shown by Hall & Smith that the fact that $n^{-2/3}$ is of larger order than $n^{-4/5}$, the corresponding order for the usual kernel estimator, is not a defect of the three estimators considered here, but rather a property of the estimation problem.

Appendix. Proof of theorem 2.2.

Since the terms of (2.5) are independent we have

$$\mathrm{var}(f_{nh}(x)) = \frac{1}{n}\mathrm{var}\left(\frac{-2}{\pi h^2}\int_0^{X_i}\frac{1}{\sqrt{X_i^2 - u^2}}K'\left(\frac{x-u}{h}\right)du\right) \leq$$

$$\frac{1}{n}E\left(\frac{-2}{\pi h^2}\int_0^{X_i}\frac{1}{\sqrt{X_i^2 - u^2}}K'\left(\frac{x-u}{h}\right)du\right)^2,$$

so it suffices to show the following bound for x>0 and h small enough

$$E\left(\frac{-2}{\pi h^2}\int_0^{X_i}\frac{1}{\sqrt{X_i^2 - u^2}}K'\left(\frac{x-u}{h}\right)du\right)^2 \leq \frac{g(x)}{x}O\left(\frac{-\log(h)}{h^2}\right)+\frac{1}{x}o\left(\frac{-\log(h)}{h^2}\right) \qquad (7.1)$$

By two successive substitutions v=(u-x)/h and z=(t-x)/h we get for h<x

$$E\left(\frac{-2}{\pi h^2}\int_0^{X_i}\frac{1}{\sqrt{X_i^2 - u^2}}K'\left(\frac{x-u}{h}\right)du\right)^2 =$$

$$\frac{4}{\pi^2 h^4}\int_0^{\infty}\left(\int_0^{t}\frac{1}{\sqrt{t^2 - u^2}}K'\left(\frac{x-u}{h}\right)du\right)^2 g(t)dt =$$

$$\frac{4}{\pi^2 h^2}\int_0^{\infty}\left(\int_{-x/h}^{(t-x)/h}\frac{1}{\sqrt{t^2 - (x+hv)^2}}K'(v)dv\right)^2 g(t)dt =$$

$$\frac{4}{\pi^2 h}\int_{-x/h}^{\infty}\left(\int_{-x/h}^{z}\frac{1}{\sqrt{2hx(z-v) + h^2(z^2-v^2)}}K'(v)dv\right)^2 g(x+hz)dz =$$

$$\frac{2}{\pi^2 h^2}\int_{-1}^{\infty}\left(\int_{-1}^{z}\frac{1}{\sqrt{z-v}}\,\frac{1}{\sqrt{x + \frac{1}{2}h(z+v)}}K'(v)dv\right)^2 g(x+hz)dz.$$

Next write this integral as the sum of two terms

$$\frac{2}{\pi^2 h^2}\int_{-1}^{1}\left(\int_{-1}^{z}\frac{1}{\sqrt{z-v}}\,\frac{1}{\sqrt{x + \frac{1}{2}h(z+v)}}K'(v)dv\right)^2 g(x+hz)dz \qquad (7.2)$$

and

$$\frac{2}{\pi^2 h^2}\int_{1}^{\infty}\left(\int_{-1}^{z}\frac{1}{\sqrt{z-v}}\,\frac{1}{\sqrt{x + \frac{1}{2}h(z+v)}}K'(v)dv\right)^2 g(x+hz)dz, \qquad (7.3)$$

which we shall treat separately.

By the conditions on f relation (1.1) implies that g is a bounded density and thus it is readily shown that the term (7.2) is bounded by a constant times $1/(xh^2)$ for h small enough and so it satisfies (7.1). For the term (7.3) the argument is more involved. For $z \geq 1$ we have for some positive constants c and c'

$$\left| \int_{-1}^{z} \frac{1}{\sqrt{z-v}} \frac{1}{\sqrt{x + \frac{1}{2}h(z+v)}} K'(v)dv \right| \leq$$

$$\frac{c}{\sqrt{x}} \int_{-1}^{1} \frac{1}{\sqrt{z-v}} dv =$$

$$\frac{2c}{\sqrt{x}} (\sqrt{z+1} - \sqrt{z-1}) \leq$$

$$\frac{c'}{\sqrt{x}} \frac{1}{\sqrt{z}},$$

since it follows from the mean value theorem that for all $z \geq 1$ we have

$$\sqrt{z+1} - \sqrt{z-1} = \sqrt{z} \left(\sqrt{1 + \frac{1}{z}} - \sqrt{1 - \frac{1}{z}} \right) \leq \frac{c''}{\sqrt{z}},$$

for some positive constant c". By this bound the term (7.3) is dominated by

$$\frac{2c'^2}{x\pi^2 h^2} \int_{1}^{\infty} \frac{1}{z} g(x+hz)dz = \frac{2c'^2}{x\pi^2 h^2} \int_{h}^{R} \frac{1}{y} g(x+y)dy.$$

Rewriting the integral we get

$$\int_{h}^{R} \frac{1}{y} g(x+y)dy = - g(x) \log(h) + \int_{h}^{1} \frac{1}{y} (g(x+y) - g(x))dy + \int_{1}^{R} \frac{1}{y} g(x+y)dy,$$

and consequently

$$\left| \int_{h}^{R} \frac{1}{y} g(x+y)dy + g(x) \log(h) \right| \leq \int_{h}^{1} \frac{1}{y} |g(x+y) - g(x)|dy + \int_{1}^{R} \frac{1}{y} g(x+y)dy.$$

By our conditions on f the density g is bounded and continuous on [0,R], so the second term is smaller than a constant independent of x. Next, using the fact that g is uniformly continuous on [0,R], we see that for every $\varepsilon>0$ there exists a $\delta>0$ such that $0<y<\delta$ implies $|g(x+y) - g(x)| \leq \varepsilon$. This gives for $h \leq \delta$

$$\frac{-1}{\log(h)}\int_h^1 \frac{1}{y}|g(x+y) - g(x)|dy =$$

$$\frac{-1}{\log(h)}\int_h^\delta \frac{1}{y}|g(x+y) - g(x)|dy + \frac{-1}{\log(h)}\int_\delta^1 \frac{1}{y}|g(x+y) - g(x)|dy \leq$$

$$\varepsilon\frac{\log(\delta) - \log(h)}{-\log(h)} + \frac{2}{-\delta\log(h)}.$$

Since this bound is smaller than 2ε for h small enough we have now shown

$$\int_h^R \frac{1}{y}g(x+y)dy = -g(x)\log(h) + o(-\log(h)),$$

which implies that the term (7.3) also satisfies (7.1). □

References.

Anderssen, R.S. and A.J. Jakeman (1975), Abel type integral equations in stereology, J. Microsc. 105, 135-153.

Devroye, L. (1987), A Course in Density Estimation, Birkhäuser, Boston.

Devroye, L. and L. Györfi (1985), Nonparametric Density Estimation, The L_1 view, Wiley, New York.

Es, A.J. van (1988), Aspects of Nonparametric Density Estimation, Ph.D. Thesis, University of Amsterdam, In preparation.

Hall, P. and J.S. Marron (1987a), Extent to which least squares cross-validation minimizes integrated squared error in nonparametric density estimation, Probab. Th. Rel. Fields 74, 567-581.

Hall, P. and J.S. Marron (1987b), On the amount of noise inherent in bandwidth selection for a kernel density estimator, Ann. Stat. 15, 163-181.

Hall, P. and R.L. Smith (1984), Unfolding a nonparametric density estimate, Technical Report, Imperial College, London.

Hall, P. and R.L. Smith (1988), The kernel method for unfolding sphere distributions, J. Comp. Phys. 74, 409-421.

Nychka, D., Wahba, G. Goldfarb, S. and T. Pugh (1984), Cross-validated spline methods for the estimation of three-dimensional tumor size distributions from observations on two dimensional cross sections, J. Amer. Statist. Ass. 79, 832-846.

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

Ripley, B.D. (1987), Stochastic Simulation, Wiley, New York.

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

Stoyan, D. , Kendall, W.S. and J. Mecke (1987), Stochastic Geometry and Its Applications, Akademie-Verlag, Berlin.

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

Watson, G.S. (1971), Estimating functionals of particle size distributions, Biometrika 58, 483-490.

Wicksell, S.D. (1925), The corpuscle problem, Part I, Biometrika 17, 84-99.

Wilson, J.D. (1987), A smoothed EM algorithm for the solution of Wicksell's corpuscle problem, Technical Report, University of Bath.