



Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

A.W. Hoogendoorn

Estimating the weight undersize distribution for
the Wicksell problem

Department of Mathematical Statistics

Report MS-R8811

December

Bibliotheek
Centrum voor Wiskunde en Informatica
Amsterdam

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 Research (N.W.O.).

Copyright © Stichting Mathematisch Centrum, Amsterdam

Estimating the Weight Undersize Distribution for the Wicksell Problem

Adriaan W. Hoogendoorn

Centrum voor Wiskunde en Informatica,
Kruislaan 413, 1098 SJ Amsterdam

Suppose that spherical particles are contained in some opaque body (so they cannot be observed), and one is interested in the distribution of their diameters. Examples of such situations come from geology and biology, if one is interested in oil-drops that are contained in a piece of rock, or carcinomas in mouse-livers. In these situations often the only information one can get is by making a cross-section of that body and measuring the circle profiles. The problem to estimate the distribution of the sphere-diameters from the observed circle-diameters is known as Wicksell's corpuscle problem.

In this paper a review will be given of methods that have been applied to solve this problem. Also a study is made of an estimator for the Weight Undersize Distribution that was proposed by ENGELS (1986) by whom we came to this subject.

Key words and Phrases: particle size distribution, smoothing, stereology, weight undersize distribution.

1980 Mathematical Subject Classification: 62G05.

1. REVIEW OF THE WICKSELL PROBLEM

1.1 Introduction

WICKSELL (1925) derived a relationship between the distribution-functions F (of sphere-diameters) and G (of circle-diameters) under the assumption that the centres of the spheres are points of a Poisson process with intensity so low, that the spheres rarely intersect. Under this assumption two considerations lead to the relationship. First: the probability that a sphere is cut by a random section is proportional to its diameter, and second: conditional on the fact that a sphere is intersected, the distribution of the distance between the plane and the centre of the sphere is uniform. These considerations lead to the formula:

$$(1) \quad G(y) = 1 - \frac{1}{\mu y} \int_0^{\infty} \sqrt{x^2 - y^2} dF(x), \quad y > 0,$$

where

$$(2) \quad \mu := \int_0^{\infty} x dF(x)$$

is the expected sphere-diameter. Since this is an Abel-type integral equation it can easily be inverted, giving an expression of F in terms of G :

$$(3) \quad 1 - F(x) = \frac{2\mu}{\pi} \int_x^{\infty} \frac{dG(y)}{\sqrt{y^2 - x^2}},$$

where μ can be found by substitution of $x=0$. The same equations hold if we spoke about radii instead of diameters. Later it was proved that these equations still hold if a more general model is assumed in which the sphere-centres and -diameters form a marked point process in \mathbb{R}^3 which is stationary under translations (see BADDELEY, 1982). The sample (from one section) is no longer independent; hopefully this will only have a small effect on the results.

That the problem is not solved by having obtained formula (3) is because this relationship holds for the *true* distribution functions (we have nothing more but a sample from G) and because the inversion is numerically ill-behaved. In this sense the Wicksell problem is called an ill-posed inverse problem, meaning that small changes in G may lead to substantial changes in the corresponding F . Another important notion is that not all G 's can occur. Because equation (1) always can be differentiated, it follows that G will have a density

$$(4) \quad g(y) = \frac{y}{\mu} \int_y^{\infty} \frac{dF(x)}{\sqrt{x^2 - y^2}}$$

no matter the character of F . Since g can be written as

$$g(y) = \int_{x=0}^{\infty} \frac{y/x}{\sqrt{x^2 - y^2}} 1_{[0,x)}(y) d\tilde{F}(x),$$

where \tilde{F} is the distribution function of the sphere-diameters conditional on the sphere being intersected defined by $\tilde{F}(x) = xF(x)/\mu$, it follows that g is a mixture over $X=x$ of density-functions k , where

$$k(y;x) = \frac{y/x}{\sqrt{x^2 - y^2}} 1_{[0,x)}(y),$$

and where X follows distribution \tilde{F} (in fact there is a convolution structure behind this relationship, which will be made explicit in section 1.4). These observations may explain why estimating F by substitution of the empirical distribution-function of G into (3) gives a very unsatisfactory estimator, which is non-monotone, takes arbitrarily large negative values, and is unstable under small changes in the data. Regularization, finding a solution that is consistent both with the data and with prior assumptions, should help in this situation (see TITTERINGTON, 1985 and O'SULLIVAN, 1986). A prior assumption that is

mainly made is that the solution must be smooth. Smoothing will therefore be a key-word in solving this problem.

So many different approaches have been used to solve the problem that categories can be made. Following COLEMAN (1987) we come to half a dozen categories mentioned and explained in sections 1.2 through 1.7.

1.2 Histogram, Finite difference estimators.

The oldest technique for finding a solution is discretizing F (binning the sphere-diameters) so that equation (1) becomes a finite sum leading to a system of linear equations which can easily be solved (WICKSELL, 1925; SALTYKOV, 1967; GOLDSMITH, 1967). The smoothing operation is based on grouping the data. The difference between the papers referred to lies in choosing the position of the histogram classes; choosing either the midpoints or upper limits of the upper limit of each class in the calculations (see TAYLOR, 1983).

1.3 Product Integration.

ANDERSEN & JAKEMAN (1975) apply a numerical method using the inverse formula (3) and smoothing the empirical distribution-function G_n by a localized Lagrange interpolation before substitution. In its simplest form this means the substitution of the frequency polygon (the distribution function with density which is constant between the observed circle-diameters). A subclass of the thus obtained estimators is MSE-consistent and has finite variance (see COLEMAN, 1987).

1.4 Kernel Estimators

TAYLOR (1983) uses a kernel-estimator

$$g_{nh}(y) := \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{y-Y_i}{h}\right)$$

(where K is some density-function) to estimate g and substitutes it into the inverse formula (3) to obtain an estimator for f . The smoothing operation here is to smear the probability mass $1/n$ assigned to an observation symmetrically over its neighbourhood before transformation. TAYLOR recommends choosing the window-width dependent not only on the density of the observations, but also on the size of the observations, taking larger window-width in sparse data-regions and near zero.

HALL & SMITH (1988) use a kernel estimator to estimate the density function g_1 of the *squared* circle-diameters, defined by $g_1(v) = (2v^{1/2})^{-1} g(v^{1/2})$, and then transform it to obtain an estimator for the density function f_1 of the *squared* sphere-diameters, defined by $f_1(u) = (2u^{1/2})^{-1} f(u^{1/2})$. The relationship between these densities is given by

$$(5) \quad g_1(v) = \frac{1}{2\mu} \int_v^{\infty} \frac{1}{\sqrt{u-v}} f_1(u) du, \quad v > 0.$$

The transformation mentioned above is found by inversion of (5), giving:

$$f_1(u) = \frac{-2\mu}{\pi} \frac{d}{du} \int_u^{\infty} \frac{1}{\sqrt{v-u}} g_1(v) dv, \quad u > 0.$$

Their motivation to use squared diameters is that the relation between the densities of the squared circle- and sphere-diameters has a convolution structure which makes them easier to work with, as can be seen by rewriting (5) as

$$g_1(v) = \int_{-\infty}^{\infty} \frac{1_{[-\infty;0]}(v-u)}{2\mu\sqrt{|v-u|}} dF_1(u), \quad v > 0.$$

Note that this is a convolution with a function $k(x) = 1_{[-\infty;0]}(x) (2\mu |x|^{1/2})^{-1}$ which is not a probability density-function; but then this relationship only holds for $v > 0$. The actual bandwidth of HALL & SMITH's estimator for g depends on the size of the observed diameter Y_i (compare this remark with the recommendation of TAYLOR). To see this suppose a kernel is used with support $[-1,1]$. A classical kernel-estimator smears the probability-mass assigned to an observation Y_i over an interval $[Y_i-h, Y_i+h]$, while HALL & SMITH's estimator

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

smears out the probability mass over a neighbourhood $[Y_i-h', Y_i+h']$, where

$$h' = \sqrt{Y_i^2 + h} - Y_i = \frac{h}{2Y_i} + O(h^2), \quad \text{as } h \rightarrow 0.$$

HALL & SMITH also prove a lower bound for the rate of convergence for estimating densities f_1 with f_1 and f_1' absolutely continuous on $[0, \infty)$ and f_1, f_1' and f_1'' have an essential supremum that is bounded by a fixed number. It turns out that their, TAYLOR's and the following kernel-estimator achieve this lower bound.

VAN ES & HOOGENDOORN (1988) use the notion that a kernel-density estimator can be defined as the derivative of a kernel-smoothing operation on the distribution function:

$$f_{nh}(x) := \frac{d}{dx} \int_{-\infty}^{\infty} \frac{1}{h} K\left(\frac{x-t}{h}\right) F_n(t) dt.$$

In the classical situation F_n is the empirical distribution function, but here F_n is obtained by substitution of the distribution function of the circle-diameters G_n into equation (3). The smoothing operation can be interpreted as smearing out the transformed probability mass that was assigned to an observation. VAN ES & HOOGENDOORN also give a comparison study of the three kernel-based methods. It turns out that their estimator looks very similar to TAYLOR's. This may be expected since their method is equivalent to TAYLOR's method in the following sense: regarding the right hand side of equation (3) as an operator T applied to the distribution-function G ($F = TG$), and denoting TAYLOR's smoothing operation by S , then the same solution will be found if VAN ES & HOOGENDOORN's method is applied using the smoothing operation TST^{-1} .

1.5 Splines

NYCHKA, et al. (1984) use a cross-validated spline method to estimate the density f . By a partitioning of the positive halfline into intervals $[P_i, P_{i+1})$ let the data $\{z_i\}$ be the fraction of observations from our sample in the i 'th interval ($i = 1, \dots, m$). The true fraction ζ_i , which is $\zeta_i = G(P_{i+1}) - G(P_i)$ can by (1) be written as

$$\zeta_i = L_i(f/\mu) := \int_{P_i}^{\infty} \sqrt{x^2 - P_i^2} f(x)/\mu dx - \int_{P_{i+1}}^{\infty} \sqrt{x^2 - P_{i+1}^2} f(x)/\mu dx,$$

thus defining a linear functional L_i applied to f/μ . The model becomes $z_i = L_i h + \varepsilon_i$ ($i=1, \dots, m$), where the authors assume that the $\{\varepsilon_i\}$ are independent zero-mean random variables with common variance. Estimators of f/μ are obtained by minimizing

$$\frac{1}{m} \sum_{i=1}^m (L_i h - z_i)^2 + \lambda \int_0^{\infty} (h''(x))^2 dx,$$

over functions h , where the first part forces the solution to stay close to the observations and the second part forces the solution to be smooth. The trade-off between these two effects is guided by the smoothing parameter λ ; putting λ zero means that classical interpolation is used, since no smoothness is demanded, while letting λ go to infinity leads to the use of classical linear least squares, since the second term is zero if and only if h is linear. If λ is somewhere between 0 and ∞ then it is known that the solution is a piecewise cubic polynomial (a cubic spline). NYCHKA, et al. suggest that λ should be chosen by cross-validation. Some criticism can be made. First: because the $\{\varepsilon_i\}$ are not i.i.d. random variables why not use a generalized least squares method used instead, since the covariance structure can easily be estimated. Second: why discretize in advance and then use a Penalized Least Squares method, instead of estimating the density f directly by the Maximum Penalized Likelihood (MPL) method on the pure data. The MPL-

estimator for the classical situation (where the observations come from the density under interest) is found by maximizing the functional

$$l(f) = \sum_{i=1}^n \log f(x_i) - \alpha R(f),$$

where R is a functional (sometimes called the "Roughness-penalty") that can be chosen according to prior information, and α is the smoothings-parameter. An example of a Roughness-penalty is

$$R(f) = \int_{-\infty}^{\infty} (f^{(m)}(x))^2 dx, \quad (m \geq 2),$$

where polynomials of degree $m-1$ have no penalty (are "infinitely smooth") and where it is known that the solution is a spline of degree $2m-1$. Another example is

$$R(f) = \int_{-\infty}^{\infty} \left(\frac{d^3}{dx^3} \log f(x) \right)^2 dx,$$

where a Gauss-function is infinitely smooth, and solutions are forced to take a shape which is close to it. Because in the Wicksell problem we have observations from g , but are interested in f , we substitute (4) into the functional l giving:

$$l(f) = \sum_{i=1}^n \log \left\{ \frac{y_i}{\mu} \int_{y_i}^{\infty} \frac{f(x)}{\sqrt{x^2 - y_i^2}} \right\} - \alpha R(f),$$

to be maximized over all densities f . Future research is needed to find an efficient way of (approximately) solving this equation numerically.

1.6 EMS-algorithm

WILSON (1987) uses a smoothed EM-algorithm to find a histogram-estimator for f , behaving as if the spheres are distributed according to a spatial Poisson-process with unknown intensity λ . Her method combines the EM-algorithm, an iterative procedure to find maximum-likelihood estimators in so-called "incomplete"-data problems (which will be explained later), with simple smoothing.

Given a partition of the range of possible circle-diameters into N intervals, let our data be $n_Y = \{n_Y(i), i=1, \dots, N\}$, the numbers of observed circle-diameters in each interval. From this data we want to find a histogram estimator for f ; therefore we discretize f by making a partition of its (bounded) support into M equal intervals (of length h_M) and suppose that f takes a constant value on each interval. Then the sphere intensity λf at a given diameter x in the j 'th interval ($j = 1, \dots, M$) is

$$\lambda f(x) = \frac{\varphi(j)}{h_M}$$

where $\varphi(j)$ is the expectation of $n_X(j)$, the numbers of spheres in the j 'th interval, which we cannot observe but for which we know from the original Poisson process assumption that they are Poisson distributed random variables with expectation $\varphi(j)$, and

$$\sum_{j=1}^M \varphi(j) = \lambda.$$

Let $N(i,j)$ ($i=1,\dots,N$, $j=1,\dots,M$) be the number of spheres with diameter in interval j , that are recorded as circles with diameter in interval i . Note that $n_Y(i) = \sum_{j=1}^M N(i,j)$ ($i=1,\dots,N$), but $n_X(j) \neq \sum_{i=1}^N N(i,j)$ ($j=1,\dots,M$), since not all spheres are intersected. Let further $p(i,j)$ ($i=1,\dots,N$, $j=1,\dots,M$) be the conditional probability that a sphere with diameter in interval j will be recorded as a circle with diameter in interval i . It follows that the random variables $N(i,j)$ ($i=1,\dots,N$, $j=1,\dots,M$) are Poisson distributed with mean

$$E[N(i,j)] = \varphi(j) p(i,j),$$

where the probabilities $p(i,j)$ are theoretically known. The $N(i,j)$'s are sufficient for the $\varphi(j)$'s, but unknown. To estimate the parameter vector φ (and thereby f) from our data n_Y (which is "incomplete" in the sense that the information from what sphere the circle was obtained is lost) the EM-algorithm (fully described in DEMPSTER et al. 1977) uses the known likelihood of the complete data $\text{Log}(N | \varphi)$, which is a function $F(t(N), \varphi)$ of φ and of the sufficient statistic $t(N)$ given by

$$t(N)_j = \sum_{i=1}^n N(i,j) \quad (j = 1,\dots,M)$$

in which it is linear(!). Let φ_p be the estimator for the parameter vector φ after p iterations, then this algorithm proceeds as follows: Calculate

$$t_p := E\left[\sum_{i=1}^M N(i,j) \mid n_Y, \varphi_p\right],$$

(Expectation-or E-step), and then maximize

$$F(t_p, \varphi)$$

to obtain φ_{p+1} (Maximization- or M-step). The convergence properties of the EM-algorithm are known (see WU, 1983, and VARDI et al., 1985). Since it turns out that in this situation the algorithm converges very slowly and gives very unsmooth solutions WILSON suggests adding a third step, a Smoothing- or S-

step, into the algorithm. The smoothing step consists of taking weighted averages of the $\phi_p(j)$'s with its neighbours. Though no convergence properties are proved for this EMS-algorithm, Monte Carlo simulations suggest that the method works nicely.

1.7 Non-parametric maximum likelihood

Inspired by the Wicksell problem VAN ES (1988) and VAN ES & GROENEBOOM (1988) consider the problem of recovering a distribution function from a convolution. In that situation we have observations V_1, \dots, V_n from a distribution function G which is a convolution of two other distribution functions K and F :

$$G(v) = \int_{-\infty}^{\infty} K(v-u) dF(u),$$

where we assume that K is known and that F is uniquely determined by G and K . Results are obtained showing lower bounds for a local minimax-risk for estimation of F at a fixed point x_0 for different choices of K . It turns out that the rate of convergence depends on smoothness properties of K (being worse for smooth K). As an example a nonparametric maximum likelihood estimator (NPMLE) for F is derived in the case that K is a uniform distribution. The presented distribution theory for this estimator suggests that the lower-bound is sharp. In the case of the Wicksell problem a convolution structure appears in the relationship between the squared diameters (see section 1.4, HALL & SMITH). If we assume that F has support $[0,1]$ equation (5) implies that the log-likelihood is given by

$$L(V_1, \dots, V_n) = \sum_{i=1}^n \log \left\{ \frac{1}{\gamma} \int_{V_i}^1 (u-V_i)^{-1/2} dF_1(u) \right\},$$

where $\gamma = 2\mu$ can be written as

$$\gamma = \int_0^1 \int_{V_i}^1 (v-u)^{-1/2} dF_1(v) du.$$

The situation differs from the problem above since here we have no convolution with a distribution function, and we have the extra parameter γ . The NPMLE for F_1 is a discrete distribution function, with masses $\alpha_2, \dots, \alpha_n$ at the (ordered) points V_2, \dots, V_n . This means that we maximize:

$$L(V_1, \dots, V_n) = \sum_{i=1}^{n-1} \log \left\{ \frac{1}{\gamma} \sum_{j=i+1}^n (V_j - V_i)^{-1/2} \alpha_j \right\},$$

where (defining $V_0=0$ and $\alpha_1=0$)

$$\begin{aligned}\gamma &= \sum_{i=1}^n \sum_{j=i}^n 2\alpha_j \{(V_j - V_{i-1})^{1/2} - (V_j - V_i)^{1/2}\} \\ &= 2 \sum_{j=2}^n \alpha_j V_j^{1/2}\end{aligned}$$

The maximizing vector $\alpha = (\alpha_2, \dots, \alpha_n)$ can be found using gradient projection algorithm. Though no theory is given for this estimator, a rate of convergence faster than $n^{-1/3}$ is expected, because of the peakedness of the function $K(u) = (u - V_j)^{-1/2}$. Notice that no smoothing-operation has been used here. If one is interested in the density f , a kernel estimator can be defined by differentiating the smoothed version of this NPMLE (see section 1.4, VAN ES & HOOGENDOORN).

2. ESTIMATING THE WEIGHT UNDERSIZE DISTRIBUTION.

2.1 Introduction

In this section we are interested in estimating the Weight Undersize Distribution (WUD) of the spheres at a certain point. By the WUD we mean the fraction of the volume (or weight) that is contained in spheres smaller than a certain diameter. Here we were motivated by ENGELS (1986) who needed to estimate precisely this quantity for certain practical purposes.

The formal definition of $W(z)$, the WUD in a point z , is

$$(6) \quad W(z) := \frac{1}{\mu_3} \int_0^z x^3 dF(x),$$

where μ_3 is the third moment of F . By substitution of (3) we find an expression of the WUD in terms of G (see WATSON, 1971):

$$(7) \quad W(z) = \frac{\frac{3}{2} \int_0^z y^2 dG(y) + \frac{1}{\pi z} \int \left\{ 3y^2 \sin^{-1}\left(\frac{z}{y}\right) - \frac{2z^3}{\sqrt{y^2 - z^2}} - 3z\sqrt{y^2 - z^2} \right\} dG(y)}{\frac{3}{2} \int_0^\infty y^2 dG(y)}$$

In the sequel we will frequently refer to the first, second, third and fourth term of then numerator of (7) and its empirical counterpart by considering the second integral here as the sum of three terms.

The most straightforward estimator for the WUD is to substitute the empirical distribution function G_n into equation (7). Unfortunately this estimator may take negative values and turns out to have an infinite variance; these and other properties of this estimator are discussed in section 2.2. In section 2.3 we will consider an estimator proposed by ENGELS (1986) which has a finite variance. This estimator is obtained

by smoothing the empirical distribution function locally at its singularity-point z before substitution. For convenience we will also introduce the following notation:

$$(8) \quad T_z(y) := \begin{cases} \frac{1}{\sqrt{y^2-z^2}} & y > z \\ 0 & y \leq z; \end{cases}$$

$$(9) \quad \tau_z := E_G T_z(Y) = \int_0^{\infty} T_z(y) dG(y) \\ = \frac{\pi}{2\mu} (1-F(z)), \text{ by (3).}$$

2.2. Substitution of the Empirical Distribution function.

Substitution of the empirical distribution function G_n for G gives us the following estimator for $W(z)$

$$(10) \quad W_n(z) = \frac{\frac{3}{2n} \sum_{y_i \leq z} y_i^2 + \frac{1}{\pi n} \sum_{y_i > z} \left(3y_i^2 \sin^{-1}\left(\frac{z}{y_i}\right) - \frac{2z^3}{\sqrt{y_i^2-z^2}} - 3z\sqrt{y_i^2-z^2} \right)}{\frac{3}{2n} \sum_i y_i^2}$$

In this paragraph we will consider some properties of this estimator.

(i). $W_n(z)$ may take negative values. A circle-diameter close to z may make the third term in the numerator of $W_n(z)$ and thereby the estimator itself very negative.

(ii). $W_n(z)$ is a ratio of random variables with finite expectation. Under the assumption that the sphere-diameters have a finite third moment (which is made by defining the WUD) the first term of the numerator and the denominator will have finite expectation, because the circle-diameters will have a finite second moment. (see Appendix). The expectation of the third part of the numerator has a finite expectation because by (3) we know that

$$E_G T_z(Y) = \frac{\pi}{2\mu} (1-F(z)).$$

For the first and third term of the second part of the numerator it is clear that they have a finite expectation.

(iii). $W_n(z)$ is a strongly consistent estimator. This follows now from the strong law of large numbers applied to the numerator and denominator separately.

(iv). $W_n(z)$ has an infinite variance. We consider the second moment of the third term of the numerator, and assume that $F(z) < 1$, so that the distribution of the sphere radii has a positive probability mass on (z, ∞) . It follows that on an interval $(z, z+\epsilon)$, for certain $\epsilon > 0$, the density g is positive: $g(y) \geq c_\epsilon > 0$ for $y \in (z, z+\epsilon)$. The variance of $W_n(z)$ will then be infinite since

$$E T_Z^2 = \int_z^\infty \frac{1}{y^2-z^2} dG(y) \geq c_\varepsilon \int_z^{z+\varepsilon} \frac{1}{y^2-z^2} dy = c_\varepsilon \left[\frac{1}{2z} \log \left(\frac{y-z}{y+z} \right) \right]_{y=z}^{y=z+\varepsilon} = \infty.$$

(v). *Asymptotic Normality*: WATSON (1971) shows how it is possible to prove asymptotic normality for this estimator. Under the condition that $\int_0^\infty \frac{dF(x)}{x}$ and $\int_z^\infty \frac{dF(x)}{\sqrt{x^2-z^2}}$ are bounded, it follows that:

$$\sqrt{\frac{n}{\log n}} \left(\frac{1}{n} \sum_i T_Z(Y_i) - \tau_Z \right) \rightarrow N(0, \sigma^2_T),$$

where

$$\sigma^2_T = \frac{1}{2\mu} \int_z^\infty \frac{dF(x)}{\sqrt{x^2-z^2}}.$$

We remark that this result can also be inferred from theorem 1a(i) of FELLER (1971), p 313. Notice the factor $\sqrt{n/\log n}$, lower than the usual \sqrt{n} -rate. It follows by an application of the δ -method that:

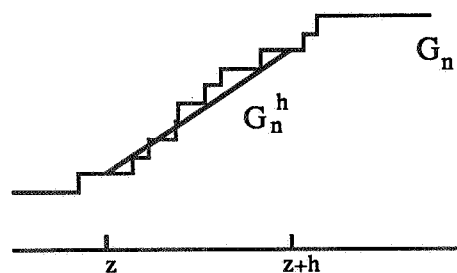
$$W_n(z) = AN \left(W(z); \frac{\log n}{n} \frac{4z^6 \sigma^2_T}{\pi^2 \beta^2} \right),$$

where

$$\beta := EGY^2 = \frac{2\mu_3}{3\mu}.$$

2.3. Smoothing the Empirical Distribution function.

By locally smoothing the empirical distribution function before substitution into (7) we hope to find an estimator for the Weight Undersize Distribution that is less sensitive to observations that are close to the point of interest. To estimate the WUD in a point z , we will smooth on the interval $[z, z+h]$ (for certain positive h), thus overcoming the singularity in T_Z . In a picture:



In formula:

$$(11) \quad G_n^h(y) := G_n(y) 1_{(0,z) \cup (z+h, \infty)}(y) + \left(G_n(z) + \frac{y-z}{h} \right) [G_n(z+h) - G_n(z)] 1_{[z, z+h]}(y)$$

We define a new estimator $W_n^h(z)$ for $W(z)$, by substitution of G_n^h into formula (7).

$$(12) \quad W_n^h(z) := \frac{\frac{3}{2} \int_0^z y^2 dG_n^h(y) + \frac{1}{\pi} \int_z^\infty \left\{ 3y^2 \sin^{-1}\left(\frac{z}{y}\right) - \frac{2z^3}{\sqrt{y^2-z^2}} - 3z\sqrt{y^2-z^2} \right\} dG_n^h(y)}{\frac{3}{2} \int_0^\infty y^2 dG_n^h(y)}$$

Later we will choose h as a function of n such that our estimator will be consistent and optimal (in MSE-sense). The setup of this paragraph will be as follows: in a first lemma something will be said about estimating T_z , the term containing the singularity. Then estimators for the numerator and the denominator of (12) will be considered in a second lemma. Some properties of the estimator $W_n^h(z)$ will finally be summarized in a theorem.

Define as an estimator for $\tau_z = \int_0^\infty T_z(y) dG(y)$,

$$(13) \quad \tau_z^h := \int_0^\infty T_z(y) dG_n^h(y) = \frac{G_n(z+h) - G_n(z)}{h} \int_z^{z+h} T_z(y) dy + \frac{1}{n} \sum_{y_i > z+h} T_z(y_i)$$

LEMMA 2.1. *Under the assumption that G has a density g with continuous second derivative in a neighbourhood $[z, z+\epsilon]$, $\epsilon > 0$ of z , we can make the following statements about the bias and variance of τ_z^h :*

- (a) bias $\tau_z^h = g'(z) (2z)^{-1/2} h^{3/2} + O(h^2)$, as $h \rightarrow 0$;
- (b) var $\tau_z^h = -n^{-1} g(z) (2z)^{-1} \log h + n^{-1} O(1)$, as $h \rightarrow 0$;
- (c) For a sequence (h_n) with $h_n \rightarrow 0$ and $-n^{-1} \log h_n \rightarrow 0$ τ_z^h is a consistent estimator;
- (d) The sequence (h_n) that minimizes asymptotically the mean square error (MSE) is

$$(14) \quad h_n = \frac{1}{3} \left(\frac{g(z)}{(g'(z))^2} \right)^{1/3} n^{-1/3}$$

and in that case $MSE = \frac{1}{3} g(z) (2z)^{-1} n^{-1} \log n + O(n^{-1})$, $n \rightarrow \infty$.

PROOF bias $\tau_z^h = E\tau_z^h - \tau_z = E \int_0^\infty T_z(y) dG_n^h(y) - \int_0^\infty T_z(y) dG(y)$.

Substitution of the empirical distribution function will give an unbiased estimator, so:

$$\begin{aligned} \text{bias } \tau_z^h &= E \int_z^{z+h} T_z(y) dG_n^h(y) - \int_z^{z+h} T_z(y) dG(y) \\ &= E \int_z^{z+h} T_z(y) \frac{G_n(z+h) - G_n(z)}{h} dy - \int_z^{z+h} T_z(y) dG(y) \\ &= \frac{G(z+h) - G(z)}{h} \int_z^{z+h} T_z(y) dy - \int_z^{z+h} T_z(y) dG(y), \end{aligned}$$

by Fubini's theorem. Since we supposed a density with second derivative:

$$G(z+h) = G(z) + hg(z) + \frac{1}{2}h^2g'(z) + \frac{1}{6}h^3g''(\zeta_1), \quad \zeta_1 \in (z, z+h);$$

$$dG(y) = g(y) dy,$$

$$g(y) = g(z) + (y-z)g'(z) + \frac{1}{2}(y-z)^2g''(\zeta_2), \quad \zeta_2 \in (z, y).$$

Substitution gives:

$$\begin{aligned} \text{bias } \tau_z^h &= \left\{ g(z) + \frac{1}{2}hg'(z) + \frac{1}{6}h^2g''(\zeta_1) \right\} \int_z^{z+h} T_Z(y) dy \\ &\quad - \int_z^{z+h} T_Z(y) \left\{ g(z) + (y-z)g'(z) + \frac{1}{2}(y-z)^2g''(\zeta_2) \right\} dy \\ &= \left\{ \frac{1}{2}hg'(z) + \frac{1}{6}h^2g''(\zeta_1) \right\} \int_z^{z+h} T_Z(y) dy \\ &\quad - g'(z) \int_z^{z+h} (y-z)T_Z(y) dy + g''(\zeta_2') \int_z^{z+h} (y-z)^2T_Z(y) dy, \quad \zeta_2' \in [z, z+h]. \end{aligned}$$

The last equation holds, since we supposed that g has a continuous second derivative and we are integrating over a bounded interval. In order to study the terms $\int_z^{z+h} T_Z(y) dy$ and $\int_z^{z+h} (y-z)T_Z(y) dy$ define functions φ and ψ :

$$\varphi(\gamma) := \int_z^{z+\gamma^2} T_Z(y) dy = 2(2z)^{-1/2}\gamma + O(\gamma^2), \text{ as } \gamma \rightarrow 0,$$

since $\varphi(0) = 0$, $\varphi'(0) = 2(2z)^{-1/2}$ and $\varphi''(0) < \infty$. It follows that

$$(15) \quad \int_z^{z+h} T_Z(y) dy = \varphi(h^{1/2}) = 2(2z)^{-1/2} h^{1/2} + O(h), \text{ as } h \rightarrow 0;$$

$$\psi(\gamma) := \int_z^{z+\gamma^2} T_Z(y)(y-z) dy = 4(2z)^{-1/2} \gamma^3 + O(\gamma^4), \text{ as } \gamma \rightarrow 0,$$

since $\psi(0) = \psi'(0) = \psi^{(2)}(0) = 0$, $\psi^{(3)}(0) = 4(2z)^{-1/2}$ and $\psi^{(4)}(0) < \infty$. It follows that

$$\int_z^{z+h} T_Z(y)(y-z) dy = \psi(h^{1/2}) = 4(2z)^{-1/2} h^{3/2} + O(h^2), \text{ as } h \rightarrow 0.$$

Since $\int_z^{z+h} (y-z)^2 T_Z(y) dy$ is bounded by $h\psi(h^{1/2})$, substitution of the expansions of φ and ψ proves (a):

$$\text{bias} = -3g'(z)(2z)^{-1/2}h^{3/2} + O(h^2).$$

If h tends to zero, the variance of τ_z^h will tend to infinity. We are interested in the rate of convergence.

$$(16) \quad \begin{aligned} \text{var } \tau_z^h &= \text{var} \left(\frac{G_n(z+h) - G_n(z)}{h} \int_z^{z+h} T_Z(y) dy \right) + \text{var} \left(\frac{1}{n} \sum_{Y_i > z+h} T_Z(Y_i) \right) \\ &\quad + 2 \text{cov} \left(\frac{G_n(z+h) - G_n(z)}{h} \int_z^{z+h} T_Z(y) dy ; \frac{1}{n} \sum_{Y_i > z+h} T_Z(Y_i) \right) \end{aligned}$$

We will first study the second term of the right hand side of equation (16) and show later that the other terms are of a lower order.

$$\text{var} \left(\frac{1}{n} \sum_{Y_i > z+h} T_Z(Y_i) \right) = \frac{1}{n} \text{var} (T_Z(Y_1) 1_{(z+h, \infty)}(Y_1)).$$

For the second moment of $T_Z(Y_1) 1_{(z+h, \infty)}(Y_1)$ we can state:

$$E(T_Z^2(Y_1) 1_{(z+h, \infty)}(Y_1)) = \frac{g(z)}{2z} \log h + O(1), \text{ as } h \rightarrow 0.$$

To show this, we treat the integration-areas $(z+h, z+1)$ and $(z+1, \infty)$ separately:

$$E(T_Z^2(Y_1) 1_{(z+h, \infty)}(Y_1)) = \int_{z+h}^{z+1} T_Z^2(y) dG(y) + \int_{z+1}^{\infty} T_Z^2(y) dG(y).$$

The second integral

$$\int_{z+1}^{\infty} T_Z^2(y) dG(y) \leq T_Z^2(z+1) = O(1), \text{ as } h \rightarrow 0,$$

since T_Z is monotonous decreasing. For the first integral we get from the assumption that G has a density with a continuous second derivative that: $dG(y) = (g(z) + g'(z)(y-z) + g''(\zeta)(y-z)^2)dy$, where $\zeta \in (z, z+h)$, so that for certain $\zeta' \in (z, z+h)$:

$$\begin{aligned} \int_{z+h}^{z+1} T_Z^2(y) dG(y) &= g(z) \int_{z+h}^{z+1} \frac{1}{y^2-z^2} dy + g'(z) \int_{z+h}^{z+1} \frac{y-z}{y^2-z^2} dy + g''(\zeta) \int_{z+h}^{z+1} \frac{(y-z)^2}{y^2-z^2} dy \\ &= -\frac{g(z)}{2z} \log h + O(1), \text{ as } h \rightarrow 0. \end{aligned}$$

As the expectation of $T_Z(Y_1)$ is bounded for all $h > 0$:

$$\text{var} \left(\frac{1}{n} \sum_{Y_i > z+h} T_Z(Y_i) \right) = -n^{-1} \frac{g(z)}{2z} \log h + n^{-1} O(1), \text{ as } h \rightarrow 0.$$

The first term of (16):

$$\begin{aligned} \text{var} \left(\frac{G_n(z+h) - G_n(z)}{h} \int_z^{z+h} T_Z(y) dy \right) &= \frac{1}{h^2} \left(\int_z^{z+h} T_Z(y) dy \right)^2 \text{var}(G_n(z+h) - G_n(z)) \\ &= \frac{1}{h^2} \left(\int_z^{z+h} T_Z(y) dy \right)^2 n^{-1} p(1-p) \\ &= n^{-1} 2z^{-1} g(z) + n^{-1} O(h) = n^{-1} O(1), \text{ as } h \rightarrow 0, \end{aligned}$$

because of (15) and because $p = G(z+h) - G(z)$ can be written as: $p = hg(z) + O(h^2)$. The statement (b) now follows from the Cauchy-Schwartz inequality. The consistency (c) follows from (a) and (b). The optimal choice of h_n follows from minimizing the leading terms of the MSE :

$$(-3g'(z))^2 (2z)^{-1} h^3 - n^{-1}g(z)(2z)^{-1}\log h.$$

Putting the derivative (with respect to h) zero gives: $27(g'(z))^2 (2z)^{-1}h^2 = n^{-1}g(z)(2z)^{-1}h^{-1}$, with solution:

$$h_n = \frac{1}{3} \left(\frac{g(z)}{(g'(z))^2} \right)^{1/3} n^{-1/3}.$$

Substitution gives:

$$\text{MSE} = \frac{1}{3} g(z) (2z)^{-1} n^{-1} \log n + O(n^{-1}), \quad n \rightarrow \infty. \quad \square$$

In order to study properties of $W_n^h(z)$, we will consider the numerators and the denominators of (7) and (12) separately, writing

$$(17) \quad W(z) =: \frac{\alpha}{\beta},$$

and

$$(18) \quad W_n^h(z) =: \frac{A_n^h}{B_n^h}.$$

LEMMA 2.2. *Under the assumption of lemma 1 together with $E_{\mathbb{F}}X^5 < \infty$, the following statements about A_n^h and B_n^h as estimators for α and β as defined in (17) and (18) hold for $n \rightarrow \infty$ and $h \rightarrow 0$:*

- (a) bias $B_n^h = O(h^2)$;
- (b) var $B_n^h = O(n^{-1})$;
- (c) bias $A_n^h = g'(z) (2z)^{-1/2} h^{3/2} + O(h^2)$;
- (d) var $A_n^h = -n^{-1} g(z) (2z)^{-1} \log h + O(n^{-1})$.

PROOF $\text{bias } B_n^h = \frac{G(z+h)-G(z)}{h} \int_z^{z+h} y^2 dy - \int_z^{z+h} y^2 dG(y),$

by the same arguments as used in the proof of lemma 1a. Since

$$\int_z^{z+h} y^2 dy = O(h^2) \text{ as } h \rightarrow 0,$$

a Taylor-expansion of $G(z+h)$ and $G(y)$ around z will lead now to (a). Under the assumption that $E_{\mathbb{F}}X^5 < \infty$, (see Appendix) var B_n^h exists and is $O(n^{-1})$, independent of h , implying (b). The first term of the numerator of W_n^h is unbiased. For the second and the fourth term it follows by the same arguments as used in proof (1a) that their bias is $O(h^2)$. Together with lemma 1 this implies (c). Under the assumption

that $E_{\mathbb{P}}X^5 < \infty$, the variance of the first, second and fourth term is $O(n^{-1})$. Lemma 1 and the Cauchy-Schwartz inequality then gives (d). \square

THEOREM 2.1. *Under the assumption that G has a density g with continuous second derivative in a neighbourhood $[z, z+\varepsilon]$, $\varepsilon > 0$ of z , and F has finite fifth moment, we can make the following statements about the estimator $W_n^h = A_n^h/B_n^h$.*

- (a) *For a sequence (h_n) with $h_n \rightarrow 0$ and $-n^{-1} \log h_n \rightarrow 0$ W_n^h is a consistent estimator.*
 (b) *The sequence (h_n) that minimizes asymptotically the mean square error (MSE) of A_n^h is :*

$$(19) \quad h_n = \frac{1}{3} \left(\frac{g(z)}{(g'(z))^2} \right)^{1/3} n^{-1/3}$$

and in that case

$$\text{MSE}(A_n^h) = \frac{1}{3} g(z) (2z)^{-1} n^{-1} \log n + O(n^{-1}), \quad n \rightarrow \infty,$$

while

$$\text{MSE}(B_n^h) = O(n^{-1}), \quad n \rightarrow \infty,$$

which is of a lower order.

PROOF. The theorem is a direct consequence of lemma 2. \square

REMARK 2.1. We would have liked to have proved: $\text{MSE}(W_n^h) = \frac{1}{3\beta^2} g(z) (2z)^{-1} n^{-1} \log n + O(n^{-1}), \quad n \rightarrow \infty$.

However this doesn't follow from (b), since B_n^h can be arbitrarily close to zero and $E[1/B_n^h] = \infty$. Nevertheless we can prove that if we choose $h = O(n^{-1/3})$ the mean square error of W_n^h is of an order equal or less than $n^{-2/3}$. Therefore we use that $|W_n^h|$ is bounded (with probability 1) by a non-random term that is $O(n^{1/6})$ as $n \rightarrow \infty$. Then it follows that

$$\begin{aligned} \text{MSE}(W_n^h) &= E \left[\frac{A_n^h}{B_n^h} - \frac{\alpha}{\beta} \right]^2 = E \left[\frac{A_n^h - \alpha}{\beta} - \frac{A_n^h(B_n^h - \beta)}{B_n^h \beta} \right]^2 \\ &\leq E \left[\frac{A_n^h - \alpha}{\beta} \right]^2 + E \left[\frac{A_n^h(B_n^h - \beta)}{B_n^h \beta} \right]^2 + 2E \left[\left| \frac{A_n^h - \alpha}{\beta} \right| \left| \frac{A_n^h(B_n^h - \beta)}{B_n^h \beta} \right| \right] \\ &\leq \frac{1}{\beta^2} (E[A_n^h - \alpha]^2 + E[B_n^h - \beta]^2) O(n^{1/3}) + 2E[|A_n^h - \alpha| |B_n^h - \beta|] O(n^{1/6}) \\ &= O(n^{-2/3}), \quad \text{as } n \rightarrow \infty. \end{aligned}$$

To prove that $|W_n^h|$ has a nonrandom upperbound which is $O(n^{1/6})$ as $n \rightarrow \infty$ (with probability 1), or (because of our choice of h) $O(h^{-1/2})$ as $h \rightarrow 0$ (w.p. 1), we will consider three situations. First: Suppose all observations $y_i \leq z$ ($i=1, \dots, n$). In that case $A_n^h = B_n^h$ and $W_n^h = 1$. Second: Suppose there is exactly one observation $y_i > z$. Then $B_n^h > z^2 n^{-1}$. Since by definition $A_n^h < B_n^h$ and $B_n^h > 0$, it follows that $W_n^h \leq 1$ (w.p. 1), so $|W_n^h|$ can only take large values if A_n^h is negative. Therefore we are interested in

finding an upperbound for $-A_n^h$. If h tends to zero then $-A_n^h$ will tend to infinity since its behaviour depends on the third term t_3 ,

$$t_3 = \int_z^{\infty} 2z^3\pi^{-1} (y^2-z^2)^{-1/2} dG_n^h(y).$$

To study this term we consider the cases $z < y_i < z+h$ and $y_i \geq z+h$ separately. If $y_i > z+h$, then $t_3 = 2z^3\pi^{-1}n^{-1} (y_i^2-z^2)^{-1/2} \leq 2z^3\pi^{-1}n^{-1} ((z+h)^2-z^2)^{-1/2} = n^{-1}O(h^{-1/2})$, as $h \rightarrow 0$. If $z < y_i \leq z+h$ then the density on the smoothing-interval is $(nh)^{-1}$, and by (15):

$$t_3 = 2z^3\pi^{-1}(nh)^{-1} \int_z^{z+h} (y^2-z^2)^{-1/2} dy = n^{-1}O(h^{-1/2}), \text{ as } h \rightarrow 0.$$

In both cases $-A_n^h$ is bounded by a term that is $n^{-1}O(h^{-1/2})$, as $h \rightarrow 0$ (w.p. 1). But then $|A_n^h|$ is bounded by a term that is $n^{-1}O(h^{-1/2})$, as $h \rightarrow 0$ (w.p. 1), and $|W_n^h|$ is bounded by a term that is $O(h^{-1/2})$, as $h \rightarrow 0$ (w.p. 1). The third situation: Suppose there are M (stochastic many) y_i 's $> z$. In that case B_n^h is bounded by $Mn^{-1}O(1)$, as $h \rightarrow 0$, and $|A_n^h|$ is bounded by $Mn^{-1}O(h^{-1/2})$, as $h \rightarrow 0$. But then $|W_n^h|$ is bounded by a nonrandom term that is $O(h^{-1/2})$, as $h \rightarrow 0$ (w.p. 1).

REMARK 2.2. W_n^h may take negative values. In practice the estimator W_n^{h*} will be used, which is W_n^h truncated to $[0,1]$. Since $B_n^h > (0 \vee A_n^h)$, W_n^{h*} can be defined as

$$W_n^{h*} := \frac{[A_n^h]}{B_n^h},$$

where $[A_n^h] := 0 \vee A_n^h$. The mean square error for this estimator

$$\begin{aligned} \text{MSE}(W_n^{h*}) &\leq E \left[\frac{[A_n^h] - \alpha}{\beta} \right]^2 + E \left[\frac{[A_n^h] B_n^h - \beta}{B_n^h \beta} \right]^2 + 2E \left[\left| \frac{[A_n^h] - \alpha}{\beta} \right| \left| \frac{[A_n^h] B_n^h - \beta}{B_n^h \beta} \right| \right] \\ &\leq \frac{1}{\beta^2} (E[[A_n^h] - \alpha]^2 + E[B_n^h - \beta]^2 + 2E[|[A_n^h] - \alpha| |B_n^h - \beta|]) \\ &\leq \frac{1}{\beta^2} (E[A_n^h - \alpha]^2 + E[B_n^h - \beta]^2 + 2E[|A_n^h - \alpha| |B_n^h - \beta|]) \\ &= \frac{1}{3\beta^2} g(z) (2z)^{-1} n^{-1} \log n + O(n^{-1}), \quad n \rightarrow \infty \end{aligned}$$

REMARK 2.3. The optimal order $n^{-1/3}$ for the bandwidth h_n is the same order as found for the classical histogram-estimator. To be more precise: Let g be a density with bounded support and continuous derivatives up to order three (except at the endpoints), then for general x (not a midpoint of an interval) the optimal bandwidth h_n^* in MSE-sense is:

$$h_n^* = \left(\frac{g(z)}{4(g'(z))^2} \right)^{1/3} n^{-1/3} = \frac{3}{4^{1/3}} h_n,$$

(see TAPIA & THOMPSON, 1978). Still the situations are very different: the optimal MSE for histogram-estimators is $O(n^{-2/3})$, but for our estimator W_n^h it is $O(n^{-1} \log n)$.

REMARK 2.4. Since the optimal bandwidth depends on the unknown density g , it is not possible to determine it. But because of remark 1 we may use a similar rule to that often suggested for roughly optimal bandwidth choice of histogram estimators: choose h twice the interquartile-range of G_n divided by the cube root of n (see FREEDMAN & DIACONIS, 1981). Reducing this by the same factor $4^{1/3}/3$ leads to the suggestion:

$$h_n = \frac{6 \text{ interquartile-range}}{(4n)^{1/3}}$$

REMARK 2.5. A popular technique for measuring how accurate an estimator is, is the bootstrap method. It is a general method that uses numerical analysis (actually computer simulations) instead of theoretical analysis. The bootstrap method has been useful in many cases where usual asymptotic distribution theory is too complicated to be applied (see EFRON & TIBSHIRANI, 1986).

It has been proved that the bootstrap method works in many cases where the estimators have a convergence rate of \sqrt{n} . The question is how does the method behave if it is applied on the estimators W_n (based on substitution of the empirical distribution function) and W_n^h (based on substitution of a locally smoothed estimator). ATHREYA (1987a) showed that the bootstrap method fails if the method is applied to the mean of observations that have an infinite variance and belong to the domain of attraction of a stable law of order α , where $0 < \alpha < 2$. However in the case of a normal limit ($\alpha = 2$) the bootstrap method works (ATHREYA 1987b), even when the observations have an infinite variance. This latter result implies that the bootstrap "works" for estimating the distribution of

$$(20) \quad \sigma_T^{-1} \sqrt{\frac{n}{\log n}} \left(\frac{1}{n} \sum_i T_Z(Y_i) - \tau_Z \right)$$

(in the weak sense of being asymptotically valid in probability). The bootstrap version of (20) is given by

$$\frac{\sqrt{n}}{s} \left(\frac{1}{n} \sum_i T_Z(Y_i^*) - \frac{1}{n} \sum_i T_Z(Y_i) \right)$$

where s^2 denotes the sample variance of the $T_Z(Y_i)$'s, and $(Y_1^* \dots Y_n^*)$ is a random sample (with replacement) from the empirical df G_n of the Y_i 's. Since $P(T_Z(Y_i) \leq x) = G(z) + 1 - G(\sqrt{z^2 + x^{-2}}) = 1 + O(x^{-2})$, as $x \rightarrow \infty$, it can easily be verified from Theorem 1 of KLASS & TEICHER (1977) that $\frac{\sqrt{n}}{s}$ is of order $\sqrt{\frac{n}{\log n}}$ in probability. This, however, seems to imply that the bootstrap will also work (in the weak sense of being asymptotically valid in probability) in estimating the distribution of

$$(21) \quad \sigma_T^{-1} \sqrt{\frac{n}{\log n}} (W_n(z) - W(z)).$$

The bootstrap version of (21) is given by

$$\frac{\sqrt{n}}{s} (W_n^*(z) - W_n(z)),$$

where $W_n^*(z)$ is obtained from $W_n(z)$ by simply replacing the Y_i 's by the Y_i^* 's in formula (10). The reason that the bootstrap works here is that the delta-method argument used in 2.2.(v) is again applicable.

3. APPENDIX

PROPOSITION 3.1. $E_G Y^{2n}$ exists under the assumption that $E_F X^{2n+1}$ exists.

PROOF: Substitution of Wicksell's formula (1) into the definition gives:

$$\begin{aligned} E_G Y^{2n} &= \int_0^\infty y^{2n} dG(y) = \int_0^\infty \int_y^\infty \frac{y^{2n+1}}{\mu} \frac{f(x)}{\sqrt{x^2-y^2}} dx dy \\ &= \frac{1}{\mu} \int_0^\infty f(x) \int_0^x y^{2n+1} \frac{1}{\sqrt{x^2-y^2}} dy dx, \end{aligned}$$

by Fubini's theorem,

$$\begin{aligned} &= \frac{1}{\mu} \int_0^\infty x^{2n+1} f(x) \left\{ \int_0^x (y/x)^{2n+1} \frac{1}{\sqrt{1-(y/x)^2}} 1/x dy \right\} dx \\ &= \frac{1}{\mu} \int_0^\infty x^{2n+1} f(x) \left(\int_0^1 \frac{u^{2n+1}}{\sqrt{1-u^2}} du \right) dx \\ &= \prod_{k=1}^n \left(\frac{2k}{2k-1} \right) \frac{1}{\mu} \int_0^\infty x^{2n+1} f(x) dx. \end{aligned}$$

ACKNOWLEDGEMENTS

This research was done in a project specially created for civil services (TEG). The author wants to thank Richard Gill who supervised this research, Huub Engels (by whom we came to this subject), Roelof Helmers for his special help on bootstrap matters, and Arnold Dekkers and Bert van Es for their inspiring and very helpful remarks.

REFERENCES

- ANDERSSON, R.S. & JAKEMAN, A.J. (1975) Abel type integral equations in stereology. *J. Microsc.* **105**, 135-153.
- ATHREYA, K.B. (1987a) Bootstrap of the mean in the infinite variance case. *Ann. Statist.* **15**, 723-731.

- ATHREYA, K.B. (1987b) *Bootstrap of the mean in the infinite variance case*. Proc. Ist World Congress Bernoulli Society, Math. Stat., Vol. 2, VNU Science Press, 95-98.
- BADDELEY, A. (1982) Stochastic geometry: an introduction and reading-list. *Int. Statist. Rev.* **50**, 179-193.
- COLEMAN, R. (1987) *Extended version of a paper delivered by invitation at the 4th International Conference in Stereology and Stochastic Geometry at Haus der Universität, Bern, Switzerland, 12th-15th September 1987*.
- DEMPSTER, A.P., LAIRD, N.M. & RUBIN, D.B. (1977) Maximum likelihood from incomplete data via the EM Algorithm (with discussion) *J. Roy. Statist. Soc. Ser. B* **39**, 1-38.
- EFRON, B. & TIBSHIRANI, R. (1986) Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science* **1**, 54-77.
- ENGELS, H.J.M. (1986) *A robust estimation for the determination of the weight undersize distribution*. Technical Report, AKZO-company.
- ES VAN, A.J. (1988) *Aspects of nonparametric density estimation*, Ph.D. thesis, Universiteit van Amsterdam.
- ES VAN, A.J. & GROENEBOOM P. (1988) *Recovering a distribution from a convolution*, to appear.
- ES VAN, A.J. & HOOGENDOORN, A.W. (1988) *A kernel approach to estimation of the sphere radius density in Wicksell's corpuscle problem*. Report MS-R8809, Centrum voor Wiskunde en Informatica, Amsterdam.
- FREEDMAN, D. & DIACONIS, P. (1981) On the histogram as a density estimator: L_2 -theory. *Zeitschr. Wahrscheinlichkeitstheorie* **57**, 453-476.
- GOLDSMITH, P.L. (1967) The calculation of true particle size distribution from the sizes observed in a thin slice. *Brit. J. appl. Phys.* **18**, 813.
- HALL, P. & SMITH, R.L. (1988) The kernel method for unfolding sphere distributions. *J. Comput. Phys.* **74**, 409-421.
- KLASS, M. & TEICHER, H. (1977), Iterated logarithm laws for asymmetric random variables barely with or without finite mean. *Ann. Probab.* **6**, 861-874.
- NYCHKA, D., WAHBA, G., GOLDFARB, S. & PUGH, T. (1984) Cross-validated spline methods for the estimation of three-dimensional tumor size distributions from observations on two dimensional cross sections. *J. Am. Statist. Assoc.* **79**, 832-846.
- O'SULLIVAN, F. (1986) A statistical perspective on ill-posed problems. *Statistical Science* **1**, 502-527.
- SALTYKOV, S.A. (1967) The determination of the size distribution of particles in an opaque material from measurement of the size distribution of their sections. *In: Stereology* (ed. by Elias, H.) Springer, New York.
- STOYAN, D., KENDALL W.S. & MECKE, J. (1987) *Stochastic geometry and its applications*. Akademie-Verlag, Berlin.

- TAPIA, R.A. & THOMPSON, J.R. (1978) *Nonparametric Probability Density Estimation*. John Hopkins University Press, Baltimore.
- TAYLOR, C.C. (1983) A new method for unfolding sphere size distributions. *J. Microsc.* **132**, 57-66.
- TITTERINGTON, D.M. (1985) Common structure of smoothing techniques in statistics. *Int. Statist. Rev.* **53**, 141-170.
- VARDI, Y., SHEPP, L.A. & KAUFMANN, L. (1985) A statistical model for positron emission tomography. *J. Amer. Statist. Ass.* **80**, 8-37.
- 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.
- WILSON, J.D. (1987) *A smoothed EM algorithm for the solution of Wicksell's corpuscle problem*. Technical Report, University of Bath.
- WU, C.F.J. (1983) On the convergence properties of the EM algorithm. *Ann. Statist.* **11**, 95-103.

