Methods for Density Estimation in Thick-Slice Versions of Wicksell's Problem

Andrey FEUERVERGER and Peter HALL

A new, implicit method is suggested for density estimation in inverse problems, where data are drawn not from the target distribution, but rather from its image under a transformation. The approach that we propose produces density estimators that are themselves densities, without the negativity problems known to plague more explicit inversion techniques. We also suggest a general empirical approach to selecting the smoothing parameter so as to optimize performance in the context of the target distribution, rather than its image after the transformation. We apply the new methods, and competing techniques, to a thick-section Wicksell-type problem, using data on the radii of nerve terminals from the electric organ of the electric ray *Torpedo marmorata*. It is shown that statistical properties of estimators in this problem are very different from those for the thin-slice, classical Wicksell problem, and so the two cases cannot be developed simply by analogy with one another.

KEY WORDS: Bandwidth; Binwidth; Frequency polygon; Histogram; Ill-posed problem; Inverse problem; Kernel methods; Radius; Sphere; Stereology; Thick slice; Thin slice; Volterra equation.

1. INTRODUCTION

In this article we address problems of "stereological unfolding" that involve estimating a probability density f when data are available only from a distribution whose density is a transformation of f, say $g = T(f)$. To recover f, it is necessary to invert the transformation, either explicitly or implicitly. Explicit back-transformation can be problematical, not least because the back-transformation of a density estimator is often not a density and, for example, can take negative values. (See, e.g., van Es and Hoogendoorn 1990 for graphical illustrations of this difficulty.) In this article we suggest an alternative, implicit approach. We take our estimator of f to lie always in the space of probability densities and to be of a specified type, such as a histogram or frequency polygon, and we use a technique related to least squares cross-validation (Bowman 1984; Rudemo 1982) to fit it to data from the distribution with density $g = T(f)$. A second application of ideas related to cross-validation then may be used to choose the smoothing parameter. The latter technique is also appropriate for choosing the smoothing parameter even when the estimator of f is computed by other means; for example, by explicit inversion of a kernel density estimator of g.

The problems that we address here are all related to the classic example of density estimation in settings involving inverse transformations, namely that of Wicksell's (1925) problem in stereology, reviewed by Hall (1988, sec. 1.9), Hoogendoorn (1992), Ripley (1981, sec. 9.4), Stoyan, Kendall, and Mecke (1987, sec. 11.4.1), and Weibel (1980). This problem involves estimating the distribution of radii of spheres, when data on radius are available only from random, thin slices (in theory, slices of zero thickness) through the three-dimensional medium containing the spheres. In this article we address not only Wicksell's original problem, but also a version involving relatively thick sections. Despite the importance of the thick-slice case to industrial and biological applications (e.g., Anderssen and Jakeman 1975a, 1975b; Fox 1988; Goldsmith 1967), it has seldom been studied in the statistics literature. As we show, the thick-slice case has features strikingly different from those of the classical Wicksell problem, and in statistical terms cannot be adequately treated by analogy with the latter. Therefore, along with our implicit approach we also study explicit density-inversion methods in the thick-slice case. We illustrate our techniques using thick-slice data on the radii of synaptic nerve terminals from the electric organ of the electric ray *Torpedo marmorata* (see Sec. 3).

The method that we present for selecting the smoothing parameter is apparently the first to be proposed for density estimation in Wicksell-type problems. Indeed, choosing the smoothing parameter would be particularly difficult in the contexts of previous treatments, as it must be selected relative to performance in the space of densities f, not in the space of images of f. For example, standard cross-validation arguments are inappropriate, because they rely on approximating the inner product of the true density and the density estimator by the average of values of the leave-one-out density estimator at sample points. This works only if the sample points are actually drawn from the true distribution, and in the context of inverse problems they are not. Even plug-in methods, which are based on an asymptotic approximation to the optimal bandwidth, are not straightforward, because they require stereological unfolding to estimate derivatives of the sampled density, not just the density itself.

We should note that there is no absolute standardization in the literature on usage of the terms "thick" and "thin" in reference to stereological inference from sections of spheres. Our usage coincides with that of Mase (1995), for example. However, because the Fredholm or type 2 Volterra equation [i.e., our equation (4)] with which formulas in both the "thick" and "thin" cases are connected

Andrey Feuerverger is Professor of Statistics, Department of Statistics, University of Toronto, Ontario, Canada M5S 3G3 (E-mail: *andrey@utstat.toronto.edu*). Peter Hall is Professor of Statistics, Centre for Mathematics and Its Applications, Australian National University, Canberra, ACT 0200, Australia (E-mail: *peter.hall@anu.edu.au*). This work was supported by grants from the National Science & Engineering Research Council, Canada. The authors are indebted to three reviewers for suggestions which have helped to improve the presentation of this article.

is sometimes called the "thin-section equation," the applied mathematics literature in particular sometimes asserts that both settings involve "thin" slices.

The case of thick slices (i.e., slices whose thickness is comparable with the sizes of particles themselves) was first considered by Bach (1967) and Goldsmith (1967), who gave formulas for the transformations and their inverses. Goldsmith (1967) noted that the inversion problem can be treated as one of solving a Volterra integral equation of the second type, and gave an approximate linear solution that may be used to compute histogram density estimators but that would not in fact yield consistency. Anderssen and Jakeman (1974) and Jakeman and Anderssen (1975) studied numerical methods for solving the Volterra equations in both the thick- and thin-slice contexts. Mecke and Stoyan (1980) reviewed and compared inversion formulas in the thick- and thin-slice cases and gave rigorous proofs of the formulas under the assumption that the distribution of sphere centers may be interpreted as a marked point process with marks equal to radii. [In analyzing the thin-slice case, Wicksell (1925) had assumed that sphere centers are points of a sparse Poisson process; see also Baddeley 1982.] For thin slices, Taylor (1983) suggested estimating the density of sphere radii by first estimating the density of radii observed in slices and then explicitly back-transforming. Hall and Smith (1988) derived asymptotic bias, variance, and mean squared error properties of Taylor's and related estimators, and showed that their mean-squared performance is optimal in a minimax sense. Van Es and Hoogendoorn (1990) proposed alternative methods to those considered by Taylor (1983) and Hall and Smith (1988). They noted that estimators of this general type (i.e., explicit estimators) suffer from problems with negativity and differ more from the true density than they do from one another. Mase (1995) addressed estimation of distributions rather than densities, but briefly considered smoothing techniques. Carroll, van Rooij, and Ruymgaart (1991) and Johnstone and Silverman (1991), among others, addressed statistical aspects of inverse problems more generally.

Section 2 describes our methods in a general setting, appropriate to either thin or thick slices, and introduces the problem of density estimation for thick-slice data. Section 3 describes the numerical properties of our methods, through application to both real and simulated data. Section 4 summarizes theoretical properties for which the Appendix sketches a technical analysis.

2. METHODOLOGY

2.1 Implicit Inversion in the General Case

Let $\mathcal{F}$ denote a class of bounded probability densities, and let $\mathcal{G}$ be the image of $\mathcal{F}$ under a transformation T. Using data from the distribution with density $g = T(f) \in \mathcal{G}$, we wish to estimate $f \in \mathcal{F}$. To achieve this goal, we propose taking an element $\bar{f}$ of a flexible class $\bar{\mathcal{F}}$ of potential f's, such as histograms or frequency polygons. Each $\bar{f} \in \bar{\mathcal{F}}$ depends on a smoothing parameter, h say, as well as on other quantities, such as bin heights when $\bar{\mathcal{F}}$ is the class of histograms. The class $\bar{\mathcal{F}}$ is related to a sieve, and our techniques have connections to Grenander's (1981) method of sieves.

For example, we might take $\bar{\mathcal{F}}$ to be a set of histograms $\bar{f}$ having binwidth h, bins $\mathcal{B}_j$ for $1 \le j \le k$, and height t_j on the jth bin. Alternatively, $\bar{f}$ could be the linear interpolant (at bin centers) of such an estimator—that is, a frequency polygon.

Our algorithm for selecting an element of $\bar{\mathcal{F}}$ empirically is as follows. First, define a least squares goodness-of-fit criterion $\gamma(\bar{f}, f)$ to characterize the closeness of elements of $\bar{\mathcal{F}}$ to the true f. Next, develop an empirical approximation $\hat{\gamma}$ to γ, for a given value of h. Then, for a given value of h, use $\hat{\gamma}$ to select a data-dependent member of $\bar{\mathcal{F}}$ that is close to the true f. Finally, use a version of least squares cross-validation to choose h empirically.

More specifically, given a bounded, nonnegative weight function w, we define the goodness-of-fit criterion function

$$\gamma(\bar{f}, f) = \int (\bar{f} - f)^2 w = \int \bar{f}^2 w - 2 \int \bar{f} f w + \int f^2 w.$$

Ideally, for a given value h of the smoothing parameter, we would wish to choose $\bar{f} \in \bar{\mathcal{F}}$ to minimize $\gamma(\bar{f}, f)$. In analogy to cross-validation arguments, note that the first term on the far right side does not depend on f and is readily computed, and that the last term there does not depend on $\bar{f}$ and so is unimportant to the operation of minimizing $\gamma(\bar{f}, f)$. Therefore, we need only estimate $I(\bar{f}, f) = \int \bar{f} f w$. Suppose that we have an estimator $\hat{I}(\bar{f})$ of $I(\bar{f}, f)$ based on a random sample $\mathcal{Y} = \{Y_1, \ldots, Y_n\}$ from the distribution with density g. We note here that in the case of transformations of Wicksell type, we suggest unbiased estimators of the form

$$\hat{I}(\bar{f}) = n^{-1} \sum_{i=1}^{n} L(\bar{f})(Y_i), \tag{1}$$

where L is a certain linear functional defined later. Then, using our estimator $\hat{I}(\bar{f})$ of $I(\bar{f}, f)$, we now choose $\hat{f} = \bar{f}$ to minimize

$$\hat{\gamma}(\bar{f}) \equiv \int \bar{f}^2 w - 2\hat{I}(\bar{f}).$$

We call the resulting $\hat{f}$ an "implicit inversion" estimator of f.

The final step is to select h empirically. (When it is desired to explicitly indicate dependence on the smoothing parameter h, we shall write $\bar{f}_h$ and $\hat{f}_h$ in place of $\bar{f}$ and $\hat{f}$.) Assume then that $\hat{I}$ has the form in (1) and let $\hat{I}_i$ denote the version of $\hat{I}$ that arises if we replace the n sample $\mathcal{Y}$ by the $(n-1)$ sample $\mathcal{Y}_i = \mathcal{Y} \setminus \{Y_i\}$ in the definition there; specifically,

$$\hat{I}_i(\bar{f}) = (n-1)^{-1} \sum_{j \ne i} L(\bar{f})(Y_j).$$

Put

$$\hat{\gamma}_i(\bar{f}_h) = \int \bar{f}_h^2 w - 2\hat{I}_i(\bar{f}_h),$$

and let $\hat{f}_{hi}$ be the function $\bar{f}_h \in \bar{\mathcal{F}}$ that minimizes $\hat{\gamma}_i(\bar{f}_h)$. Finally, in the spirit of cross-validation (see, e.g., Bowman 1984; Rudemo 1982), let

$$\tilde{\gamma}(h) = \int \hat{f}_h^2 w - 2n^{-1} \sum_{i=1}^{n} L(\hat{f}_{hi})(Y_i), \tag{2}$$

and choose $\hat{h}$ to minimize $\tilde{\gamma}(h)$. Our final estimator of f is $\hat{f}_{\hat{h}}$.

For insight into (2), note that if we were able to directly observe variables X_i with density f, and compute density estimators $\tilde{f}_h$ (from the full sample) and $\tilde{f}_{hi}$ (from the sample with X_i omitted), then in place of (2), we would define the cross-validation criterion by

$$\int \tilde{f}_h^2 w - 2n^{-1} \sum_{i=1}^{n} \tilde{f}_{hi}(X_i) w(X_i). \tag{3}$$

This is usually motivated through the fact that it is an almost-unbiased approximation to the first two components in the expansion $\int (\tilde{f}_h - f)^2 w = \int \tilde{f}_h^2 w - 2\int f \tilde{f}_h w + \int f^2 w$ of mean integrated squared error. By using (2) instead of (3), we are replacing the "cross-product" term $\tilde{f}_{hi}(X_i) w(X_i)$ by $L(\hat{f}_{hi})(Y_i)$, where the functional L effectively inverts the transformation that took X_i to Y_i. (It also incorporates the weight w.) This method for empirical smoothing parameter selection can be used quite generally and does not require $\hat{f}_h$ to be an implicit-inversion estimator. In fact, in Sections 3 and 4 we study properties of the method in the context of explicit inversion as well.

In both the histogram and frequency polygon cases mentioned earlier in this section, the problem of minimizing $\hat{\gamma}(\bar{f})$ reduces to one of minimizing a k-variate quadratic function of the t_j's, subject to each t_j being nonnegative. Imposing the additional condition $h \sum_j t_j = 1$ ensures that the density estimator integrates to 1. (This is true for the linear interpolant as well if we insist that the first and last bins are empty.) These constrained minimization problems may now be solved using standard quadratic programming methods. We remark that our focus here on histogram estimators is due not only to their essential simplicity, and to the fact that they lead to a computationally tractable quadratic optimization, but also to the fact that histogram estimators have a long-standing tradition in the stereological literature. Similar remarks are applicable also to frequency polygon estimators, which are introduced here primarily to handle the thin-slice case, where our implicit approach will require the estimated density to be a continuous function.

We claim that when the transformation T is the identity, our "implicit" approach to defining estimators $\hat{f}_h$ by minimizing $\hat{\gamma}(\bar{f}_h)$ often reduces to familiar, "explicit" definitions, and that in such cases our argument that h be chosen to minimize $\tilde{\gamma}(h)$ turns out to be no more than the usual cross-validation approach to smoothing parameter choice. For example, if $\bar{f}$ is the histogram with bins $\mathcal{B}_j$ of width h and with respective bin heights t_j, then $\bar{f}(x) = \sum_j I(x \in \mathcal{B}_j) t_j$ and an estimator of $I(\bar{f}, f)$ is $\hat{I}(\bar{f}) = n^{-1} \sum_i \sum_j I(Y_i \in \mathcal{B}_j) t_j$. For this choice of $\hat{I}$, the estimator $\hat{f} = \bar{f}$ that minimizes $\hat{\gamma}(\bar{f})$ is the usual histogram estimator, for which ht_j equals the proportion of data that lie in bin $\mathcal{B}_j$. Further, in the case of this example, it is also clear that $\tilde{\gamma}$ reduces to the usual cross-validation criterion for the selection of h.

2.2 Implicit Inversion in the Thick-Slice Case

In the case of Wicksell-type problems for thick slices, the data may be thought of as originating in the following way. Spheres are embedded in an opaque matrix, through which a slice of thickness 2μ is taken. For any given sphere that penetrates the slice, the largest radius of all planar sections of the spheres that lie in the slice (and are parallel to its plane) is measured. (This equals the radius of the sphere itself if the center of the sphere lies in the slice, but otherwise it is strictly less than the sphere's radius. One cannot tell from the data, however, whether the sphere center lies in the slice.) In practice, the radii measurements are generally taken from circular "shadow" profiles, formed by shining a beam (of electrons, say) through the translucent slice containing the spheres. This beam has its axis perpendicular to the plane of the slice.

Data $\mathcal{Y} = \{Y_1, \ldots, Y_n\}$ on radii arising in this way have density g, say. We may express g in terms of the density f of actual sphere radii by the formula

$$(\mu + m) g(y) = \mu f(y) + y \int_y^\infty (x^2 - y^2)^{-1/2} f(x)\, dx, \tag{4}$$

where $m = \int_{0<x<\infty} x f(x)\, dx$ equals the mean radius of a sphere (see Bach 1967, 1976; Goldsmith 1967). (Of course, f and g are both supported on the positive half-line.) Equation (4) is similar to a Fredholm or type 2 Volterra equation, albeit with the integral taken over an infinite rather than a finite interval. The inverse operator is bounded, suggesting that the inverse problem associated with (4) is less ill-posed than in the case of Wicksell's (thin-slice) problem, for example. Therefore, we expect that asymptotic properties of estimators of f found by inversion may be similar, in terms of order of magnitude, to those of relatively conventional estimators of g. Theorem 1 will bear this out.

Several equivalent inversion formulas for (4) may be found in the literature. One is $f = T^{-1} g = -CA'$, where A' denotes the derivative of the function A,

$$C = (2/\pi)^{1/2} (\mu + m)/\mu, \qquad A(x) = \int_x^\infty a(x, y) g(y)\, dy,$$

$$\begin{aligned} a(x, y) &= u[\{2\pi(y^2 - x^2)\}^{1/2}/(2\mu)], \\ u(x) &= e^{x^2/2} \int_x^\infty e^{-y^2/2}\, dy \end{aligned} \tag{5}$$

(see Jakeman 1984). Starting from this property, if α is any bounded, integrable function, then, after a change of order of integration, we may show that $\int \alpha f = C \int \beta_\alpha g$, where

$$\beta_\alpha(y) = \alpha(y) u(0) - \int_0^y a_{10}(x, y) \alpha(x)\, dx \tag{6}$$

and $a_{10}(x, y) = (\partial/\partial x) a(x, y)$. Hence any expectation with respect to the density f may be converted to an expectation

with respect to g. (The converse is also true, of course.) In particular, taking $\alpha = \bar{f}w$, where $\bar{f}$ is a nonrandom function, we see that an unbiased estimator of the quantity $I(\bar{f}, f) = \int \bar{f} f w$, based on a random sample $Y_1, \ldots, Y_n$ from the distribution with density g, is given by (1), where $L(\bar{f}) \equiv C\beta_{\bar{f}w}$ and $C = C(m, \mu)$ is as given at (5).

In practice, μ is generally known, but m often is not. By a second application of the "inner product method" used earlier, we may estimate the latter directly, without estimating f first, and then substitute this estimate for the true value in our formula for L when computing first $\hat{I}$ and then our density estimator $\hat{f}$. To this end, observe that if we take $\alpha(x) \equiv x$ in (6), then we obtain $\beta_\alpha \equiv \nu$, where

$$\nu(y) = \int_0^y a(x, y)\, dx = y \int_0^1 u[\{2\pi(1 - z^2)\}^{1/2} y/(2\mu)]\, dz.$$

Hence if X and Y have densities f and g, then $E(X) = CE\{\nu(Y)\}$, or, equivalently, $m = (2/\pi)^{1/2}(1 + \mu^{-1}m)E\{\nu(Y)\}$, giving

$$m = \frac{\mu E\{\nu(Y)\}}{(\pi/2)^{1/2}\mu - E\{\nu(Y)\}}. \tag{7}$$

Assuming only that $0 < \mu < \infty$, it may be shown that $0 \le \nu(y) \le$ const. for all y. Therefore, $\nu(Y)$ has all moments finite, and in particular, $\operatorname{var}\{\nu(Y)\} < \infty$. Hence a root-$n$–consistent estimator of m is always obtained—without any moment conditions [other than $m = E(X) < \infty$] on the distribution of X—by substituting $n^{-1}\sum_i \nu(Y_i)$ for $E\{\nu(Y)\}$ in (7).

Finally, we remark here that the fact that expectations $\int \alpha f$ with respect to f can be written instead as expectations with respect to g has previously been noted in the literature concerning Wicksell-type problems (see, e.g., Anderssen 1980; Anderssen and de Hoog 1990). Anderssen developed a general inverse-problem solving technique, built around the following formula, valid in a Hilbert space context:

$$\int \alpha f = (f, \alpha) = (T^{-1}g, \alpha) = (g, (T^{-1})^*\alpha), \tag{8}$$

where $*$ refers to the adjoint and $(\cdot, \cdot)$ is the inner product. Using this approach, the implicit method described earlier can be applied to a range of inverse problems. The same ideas can be used to develop nonparametric solutions to other inverse problems, such as those where the function of interest is a regression mean rather than a density, and to define smoothing parameter choice methods in general settings. However, because we do not have practical motivation for such applications, we do not explore them here.

In the case of inverse problems that are ill-posed, so that $(T^{-1})^*$ at (8) is unbounded, some care will be required in the choice of α, as in the thin-slice case considered later, where we require α to be differentiable and to satisfy $\alpha(0) = 0$. Alternatively, $(T^{-1})^*$ will need to be replaced by a regularized version. (A point of entry to the literature on this more general viewpoint is Ruymgaart 1993; for discussion of numerical methods in general cases, see Anderssen and de Hoog 1990.)

2.3 Implicit Inversion in the Thin-Slice Case

In the thin-slice case, the expression for g in terms of the density f may be obtained by setting $\mu = 0$ in (4):

$$g(y) = \frac{y}{m} \int_y^\infty (x^2 - y^2)^{-1/2} f(x)\, dx;$$

however, the inverse to this relation cannot be obtained from (5) by simply substituting $\mu = 0$ there. In fact, the estimator $\hat{I}(\bar{f})$ suggested in Section 2.2 is also not directly applicable if $\mu = 0$. In this case we suggest basing $\hat{I}(\bar{f})$ on classical Wicksell formulas, for which the expression $f = T^{-1}g = -CA'$ continues to hold true provided that we redefine $C = 2m/\pi$ and $a(x, y) = (y^2 - x^2)^{-1/2}$. [We continue to define A in terms of a as in (5).] The expression in (6) for β_α is also no longer applicable, as both terms on the right side will be infinite. Instead, if we assume that α is differentiable and that $\alpha(0) = 0$, then we are led, as an alternative to (6), to

$$\beta_\alpha(y) = \int_0^y a(x, y)\alpha'(x)\, dx. \tag{9}$$

Thus, restricting our attention to differentiable density estimators $\bar{f}$ for which $\bar{f}(0) = 0$, and taking $w \equiv 1$ for simplicity, we may in the present case define $L(\bar{f}) \equiv C\beta_{\bar{f}}$, where $\beta_{\bar{f}}$ is now obtained using (9). Examples of differentiable density estimators include frequency polygons, obtained by interpolating between midpoints of blocks in a histogram. (The fact that such estimators are not differentiable at block centers does not cause any difficulties.) With these changes, the implicit inversion procedure is otherwise formally the same as what we have described in the thick-slice case. [Note that the choice of $w \equiv 1$ in conjunction with the use of frequency polygons results in $\alpha'(x)$ being piecewise constant in (9).]

If m is unknown and the distribution of sphere radii X has finite variance, then m may be estimated at rate $O_p(n^{\varepsilon - 1/2})$, for all $\varepsilon > 0$; for example, by $\hat{m} = (\pi/2)(n^{-1}\sum_i Y_i^{-1})^{-1}$ (see, e.g., Ripley 1981). With this choice for $L(\bar{f})$, we may then compute $\hat{I}(\bar{f})$ as described earlier, substituting $\hat{m}$ for m in the formula for C.

2.4 Explicit Inversion in the Thick-Slice Case

An alternative approach to that suggested in Sections 2.1 and 2.2 is to simply apply the inversion formula (given in Sec. 2.2) to a density estimator $\hat{g}$, computed directly from the data $\mathcal{Y}$. For example, we might define

$$\hat{g}(y) = (nh)^{-1} \sum_{i=1}^n K\{(y - Y_i)/h\},$$

where K is a kernel function and h is a bandwidth, and take

$$\begin{aligned} \hat{f}(x) &= -C\frac{d}{dx} \int_x^\infty u[\{2\pi(y^2 - x^2)\}^{1/2}/(2\mu)]\hat{g}(y)\, dy \\ &= C_1 \left\{ \hat{g}(x) + \mu^{-1} x \int_x^\infty b(x, y)\hat{g}(y)\, dy \right\}, \end{aligned} \tag{10}$$

where $C_1 = (\mu + m)/\mu$, $b(x, y) = (y^2 - x^2)^{-1/2} u_1[\{2\pi(y^2 - x^2)\}^{1/2}/(2\mu)]$, and $u_1(x) = u'(x) = xu(x) - 1$. Details of

the numerical and theoretical properties of such explicit-inversion estimators are given in Sections 3 and 4.

The method of choosing the smoothing parameter suggested in Section 2.1 has a direct analog in the case of explicit inversion, as follows. Let $\hat{f} = \hat{f}_h$ be the explicit-inversion estimator defined in (10), let $\hat{f}_{hi}$ be its version derived from the $(n-1)$ sample $\mathcal{Y} \setminus \{Y_i\}$, let $L(\bar{f}) \equiv C\beta_{\bar{f}w}$ be defined as in Section 2.2, and, using this notation, define $\tilde{\gamma}(h)$ by (2). Then our cross-validatory approach to estimating h in the explicit inversion case amounts to choosing h to minimize $\tilde{\gamma}(h)$. If m is unknown, then we may replace it by the root-n–consistent estimator of Section 2.2 in the definition of C. Likewise, our method of choosing the smoothing parameter has a version in the thin-slice, explicit-inversion context, and when m is unknown we may replace it by the $O_p(n^{\varepsilon-1/2})$ estimator of Section 2.3.

3. NUMERICAL PROPERTIES

We implemented both the implicit and explicit inversion algorithms for the thick-slice case, together with the cross-validation methodologies described in Sections 2.1, 2.2, and 2.4 using both real and simulated data. We used the S-PLUS statistical software, version 3.4, running on a Silicon Graphics R10000 Challenge L computer (see, e.g., Becker, Chambers, and Wilks 1988). Where required, optimization of quadratic functions subject to linear (equality and inequality) constraints was carried out via the S-PLUS function "solve.QP," which links S-PLUS to Fortran subroutines that implement a Goldfarb–Idnani–type algorithm. This function, written by B. Turlach, is available via the Carnegie Mellon University "statlib" website under the file name "quadprog."

Given an arbitrary density function f for the radii of spheres in space, a random sample of observations from the density g of radii measured in a slice of thickness 2μ may be generated numerically via the following device. First, introduce an auxiliary density function p for the radii of those spheres that happen to intersect with our "slice" in space. Then clearly $p(x)$ is proportional to $(2\mu + 2x)f(x)$ so that, with $m = \int_{0<x<\infty} xf(x)\,dx$ being the mean radius of the spheres,

$$p(x) = \frac{(\mu + x)f(x)}{\int(\mu + s)f(s)\,ds} = \frac{\mu}{\mu + m}f(x) + \frac{m}{\mu + m}\,\frac{xf(x)}{m};$$

that is, $p(x)$ is a mixture of the densities $f(x)$ and $xf(x)/m$ in the proportions shown. Next, if $X = x$ is an observation from $p(x)$ and Y is the "corresponding" observation from g, then straightforward geometric reasoning shows that $Y = x$ with probability $\mu/(\mu+x)$ and $Y = x(1-U^2)^{1/2}$ with probability $x/(\mu + x)$, where U is uniformly distributed on [0, 1] and is independent of X. This procedure for generating observations from g is particularly simple if f is either the gamma distribution $G(\alpha, \beta)$ or the beta distribution $B(\alpha, \beta)$, because in those cases p will be a mixture of the $G(\alpha, \beta)$ and $G(\alpha+1, \beta)$ distributions or of the $B(\alpha, \beta)$ and $B(\alpha+1, \beta)$ distributions. Furthermore, the beta distribution, which may be rescaled from [0, 1] to $[0, \xi]$, also has the advantage of having a fixed finite support, which is convenient for presenting simulation results.

All weight functions w appearing in the least squares goodness-of-fit criteria were taken as equal to 1, to avoid introducing extraneous considerations in the simulations. We also remark that for purposes of computation—to avoid unbounded integrands—the integrals appearing in (6) and (10) were first rewritten as

$$\int_0^y a_{10}(x, y)\alpha(x)\,dx = \int_0^y \alpha(x)\,d_x a(x, y)$$

and

$$\mu^{-1}x\int_x^\infty b(x, y)\hat{g}(y)\,dy = \sqrt{\frac{2}{\pi}}\int_x^\infty \frac{x}{y}\hat{g}(y)\,d_y u\left(\frac{\sqrt{2\pi(y^2 - x^2)}}{2\mu}\right),$$

and then evaluated essentially by means of their Riemann–Steiltjes approximating sums. Note that the term (x/y) appearing in the last integrand is bounded above there by unity.

Figure 1 shows the results of a typical single simulation run from a sample of size $n = 1,000$ when the radii in three-dimensional space are assumed to have a beta $B(\alpha = 6, \beta = 4)$ distribution, rescaled onto the interval [0, 40], and the thickness of the slice is taken to be $2\mu = 15$. For this particular sample, the mean radius in the slice was found to be 21.04, whereas the mean radius of the spheres in space was estimated, via (7), to be 24.04. The true value for this distribution is 24.

Figure 1(a) gives the cross-validation curve for the implicit histogram estimator, obtained by letting the histogram range over the interval [0, 40] and requiring the bandwidths (i.e., binwidths) to divide the range into a whole number of intervals. For each binwidth h selected, the corresponding point on this cross-validation curve was determined via the following computational steps:

1. Evaluate $\hat{I}(\bar{f})$ at (1) as a linear function of the bin heights t_j using $L(\bar{f}) \equiv C\beta_{\bar{f}}$ and with $\beta_{\bar{f}}$ as defined at (6).
2. Use quadratic programming to obtain the histogram $\hat{f}_h$ that minimizes the function $\hat{\gamma}(\bar{f})$ defined in the equation following (1). The function $\hat{f}_h$ is defined by means of the bin heights t_j, which are subject to the constraints $h\sum t_j = 1$ and $t_j \geq 0$.
3. Repeat steps 1 and 2 to obtain all of the leave-one-out estimators $\hat{f}_{hi}$ that minimize the leave-one-out criterion functions $\hat{\gamma}_i(\bar{f}_h)$ defined in the equation preceding (2).
4. Finally, compute the cross-validation criterion function $\tilde{\gamma}(h)$ as defined at (2).

The cross-validation curve thus obtained is seen in this instance to have a minimum at the bandwidth $h = 5$, corresponding to a subdivision of the range [0, 40] into eight bins.

Figure 1(b) is a bar plot that shows three histograms simultaneously. Over each of the eight bin ranges indicated on the horizontal axis, three bars are drawn. In each case the height of the first of the three bars gives the height (i.e., count) of the histogram for the uncorrected data; that is,

for the radii observed in the slice. The second bar gives the expected count for the known true distribution of radii in three-dimensional space. Finally, the third bar gives the height for the histogram obtained by our implicit "correction" procedure. Although the "raw" and "corrected" histograms are, of course, subject to sampling variability, the counts in the raw histogram tend to be too low in the bins corresponding to larger radii and too high in the bins corresponding to smaller radii. This is as would be expected, owing to the downward bias that occurs for radii observed in a slice. (Note that in the first bin, the expected and "corrected" counts are both imperceptibly close to 0.)

Figure 1(c) gives the cross-validation curve (for the same data as before) for the explicit inversion procedure discussed in Section 2.4. This was based on a Gaussian kernel density estimator applied to the raw data, with standard deviation of the Gaussian playing the role of the bandwidth parameter h. For each bandwidth value h selected, the corresponding point on the cross-validation curve was determined via the following computational steps:

1. Evaluate $\hat{f}_h$ as in (10), as well as the similarly determined leave-one-out estimators $\hat{f}_{hi}$.

2. Compute the cross-validation criterion function $\tilde{\gamma}(h)$ as defined at (2) using $L(\bar{f}) \equiv C\beta_{\bar{f}}$, and with $\beta_{\bar{f}}$ defined at (6) evaluated by numerical integration.

In general, the appearance of the cross-validation curves in the explicit inversion case tends to be much smoother than in the implicit inversion case. It is seen that the minimum occurs here for the bandwidth choice $h = 3$, the standard deviation of the Gaussian kernel used in the density estimator.

Figure 1(d) shows the resulting explicit inversion density estimator using $h = 3$. The solid curve here gives the true beta density function. The density estimate for the raw data is given by the dotted curve, and the dashed curve is obtained from (10). It is seen that the corrected density is centered more appropriately (i.e., higher up) than the raw density, but of course its peakedness, relative to the true curve, is reduced slightly due to the convolution with the kernel function. In this particular instance, the corrected density does not take on negative values over any portion of its range, whereas in other instances it does do so for smaller values of x (see, e.g., Fig. 3 later). This effect did not occur in the present instance, largely because this sample was generated "correctly" from the postulated model.

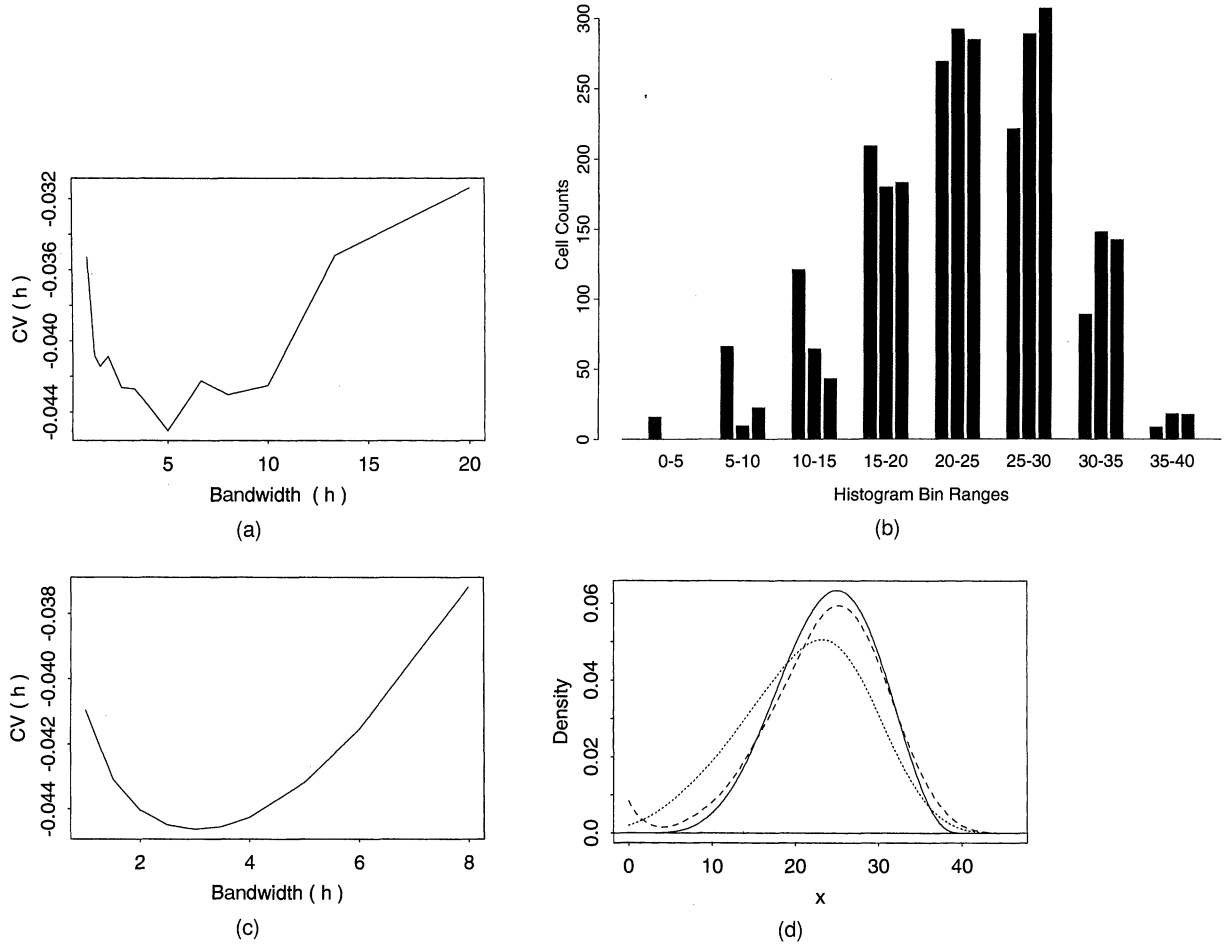

Figure 1. Analysis of a Simulated Dataset of Sample Size n = 1,000 From a Beta B (Alpha = 6, Beta = 4) Distribution Rescaled Onto the Interval [0, 40], and a Slice Thickness of 2μ = 15. (a) Cross-validation curve for the implicit histogram estimator, (b) bar plot of three histograms for observed (leftmost), theoretical (middle), and implicitly corrected (rightmost) cell counts; (c) cross-validation curve for the explicit inversion density estimator; (d) density estimate for raw data (dotted curve), the true beta density (solid curve), and the explicit inversion density estimate (dashed curve).

In real data (e.g., as in the dataset examined later), difficulties with negativity of the estimated density generally will be much more severe and occur largely as a consequence of the degradation of observations at smaller radii—that is, particles of small radius tend to not get recorded. The reader may wonder whether the "upturn" in the estimated density for x near 0 is a consequence of numerical errors. In fact, it is not; rather, it occurs due to the fact that the kernel density estimator $\hat{g}$ is not 0 at the origin. In this connection, observe that (10) is a sum of roughly commensurate terms of opposite signs, factored up by the quantity C_1. This feature generally will not arise for real data, where the left tail is usually truncated by the limited resolution of the instrumentation.

Figure 2 shows the summary results of a simulation study of the implicit histogram estimator using the same rescaled beta distribution and slice thickness indicated earlier. A chosen bandwidth of $h = 4$ (corresponding to 10 bins) was used in each of 500 Monte Carlo trials with sample size $n = 1{,}000$. (The choice $h = 4$ was among the typical bandwidths indicated by the cross-validation procedure for such data.) The figure shows the true histogram for this distribution (i.e., the expected cell counts under the true distribution). Also, in each bin, superimposed above and below the true histogram's top bar, we show the upper and lower 10th percentiles for that bin as observed in the Monte Carlo trials. In cells that do not show three distinct top bars, the lower bars are either at 0 or imperceptibly close to 0.

Figure 3 shows the summary results of a simulation study of the explicit density estimator, again using the same beta distribution, slice thickness, and sample size indicated earlier. Here, too, 500 Monte Carlo trials were conducted, and we used a fixed standard deviation of $h = 3$ for the Gaussian kernel, one of the commonly occurring bandwidths from the cross-validation. The solid curve in the figure gives the true beta density function, and the other two curves are the (pointwise) upper and lower 10th percentiles observed in the Monte Carlo trials. Note that the lower 10th percentile curve descends below 0.

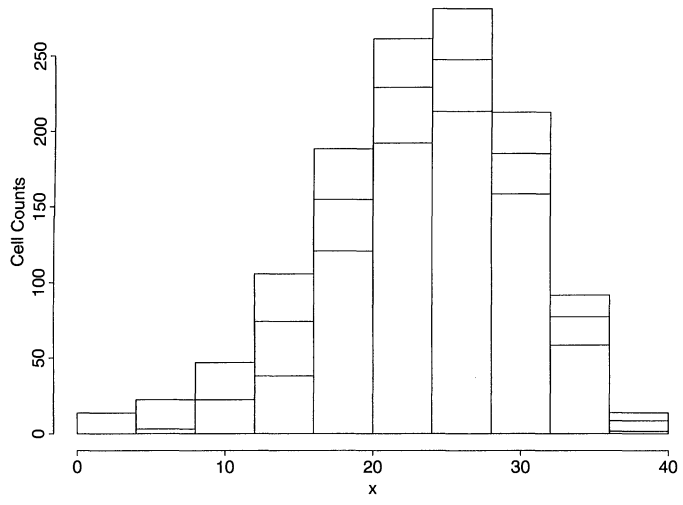

Figure 2. Results of a Simulation Study for the Implicit Histogram Estimator, With Distribution, Sample Size, and Slice Thickness as in Figure 1: True Histogram (Expected Cell Counts), With Upper and Lower 10th Percentiles for 500 Monte Carlo Trials Superimposed.

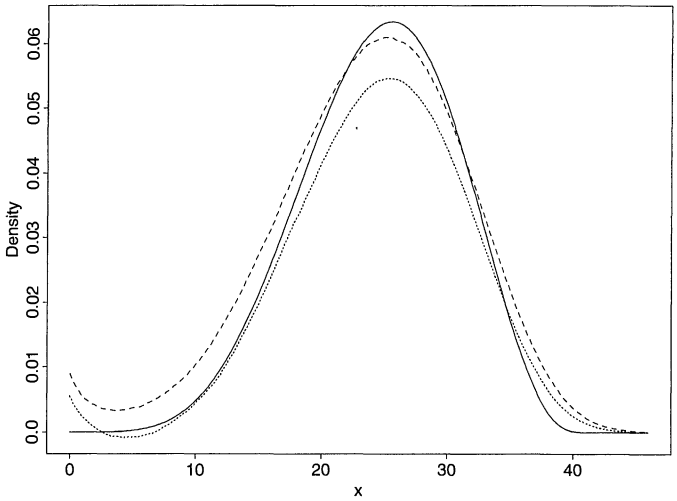

Figure 3. Results of a Simulation Study for the Explicit Density Estimator, With Distribution, Sample Size, and Slice Thickness as in Figure 1: True Density (Solid Curve) With (Pointwise) Upper and Lower 10th Percentiles of Explicit Inversion Density Estimates From 500 Monte Carlo Trials.

Figure 4 constitutes an analysis of real data that are part of a larger dataset that was very kindly supplied to us by G. Q. Fox of the Max Planck Institut für biophysikalische Chemie in Göttingen. The particular dataset examined here comprises 786 radii (in nanometers) of synaptic terminals from the electric organ of the electric ray *Torpedo marmorata,* taken from the same specimen block. Because synaptic vesicles are central to the study of synaptic function and transmission, their distributional characteristics are of significant interest. In particular, accurate comparisons of size distributions of synaptic vesicles are important in genetic mutation studies and in studies that require assessing the impact of physiological stresses or other differences arising from experimental manipulations. (For physiological and other details, see, e.g., Fox 1988 and the references therein.) A section thickness of 50 nm was used for this sample, and the observations were recorded using an electron microscope and particle analyzer. Focal depth exceeded section thickness, so the radii measured were maximum radii within the section. As is typical in data of this kind, smaller radii were lost.

These 786 observations range approximately from 17 to 53 nm, with a mean radius of 40.29 nm. The "corrected" mean radius, as computed via (7), was found to be 48.92 nm. Figure 4(a) shows the cross-validation curve for the implicit estimator using the electric ray *Torpedo* synaptic vesicle data, computed for a histogram with range (0, 55). We selected the bandwidth $h = 5$ (resulting in 11 bins), this value being in the vicinity indicated by the cross-validation curve, and allowing for a nice division for this histogram's range; the corresponding reconstruction is shown in Figure 4(b). Here again we have used a barplot method to show two histograms simultaneously. Within each bin, the leftmost bar corresponds to the raw data observed, whereas the rightmost bar corresponds to the corrected distribution.

It is seen that our procedure has the effect of shifting mass upward and, of course, that mass is never permitted to be negative. A treatment of phenomena associated with the "degradation" of observations at the smaller radii is outside the scope of this study; we note, however, that no mass is assigned to the lower ranges.

Finally, Figures 4(c) and 4(d) give the cross-validation curve for the explicit estimator, and the density estimator (10) using the value $h = 1.5$ of the kernel bandwidth, which is where the cross-validation curve takes on its minimum. The dotted curve in Figure 4(d) corresponds to the raw data, whereas the solid curve, which is seen to be "pushed upward," is the corrected density estimator. Note the substantial negative lobe of the latter—a very common occurence in real data of this type.

4. THEORETICAL PROPERTIES

4.1 Explicit Inversion

Our first result shows that, like the original kernel density estimator estimator $\hat{g}$, the back-transformed estimator $\hat{f}$ in the thick-slice case has bias of size h^2 and variance of size $(nh)^{-1}$ as $n \to \infty$.

Assume that K is a bounded, symmetric, compactly supported probability density, and put $\kappa_1 = \int K^2$ and $\kappa_2 = \int x^2 K(x)\,dx$. Suppose that $h = h(n) \to 0$ as $n \to \infty$, in such a manner that $nh \to \infty$. Assume further that f is supported on (x_1, x_2) for some $0 < x_1 < x_2 < \infty$, and has two bounded and continuous derivatives on $(0, \infty)$. (The assumption that the support of f is bounded away from the origin is realistic, because in practice, owing to physical restrictions, radii below a certain value are not observable.) Finally, suppose that the data $\mathcal{Y} = \{Y_1, \ldots, Y_n\}$ are iid and drawn from the distribution with density $g = T(f)$, where the transformation T is defined by (4), and that the estimator $\hat{g}$ is defined in terms of $\mathcal{Y}$ by (10), where m is either known or estimated root-n consistently from the data. Let (C) denote the set of conditions in this paragraph, and recall that $C_1 = (\mu + m)/\mu$.

Theorem 1. Assume conditions (C). Then

$$E\{\hat{f}(x)\} = f(x) + \frac{1}{2}h^2\kappa_2 f''(x) + o(h^2) \tag{11}$$

and

$$\text{var } \hat{f}(x) = (nh)^{-1}C_1^2\kappa_1 g(x) + o\{(nh)^{-1}\} \tag{12}$$

uniformly in $x \in (0, \infty)$ as $n \to \infty$. Furthermore, at each point in its argument, $\hat{f}$ is asymptotically normally distributed as $\mathrm{N}(E\hat{f}, \text{var } \hat{f})$.

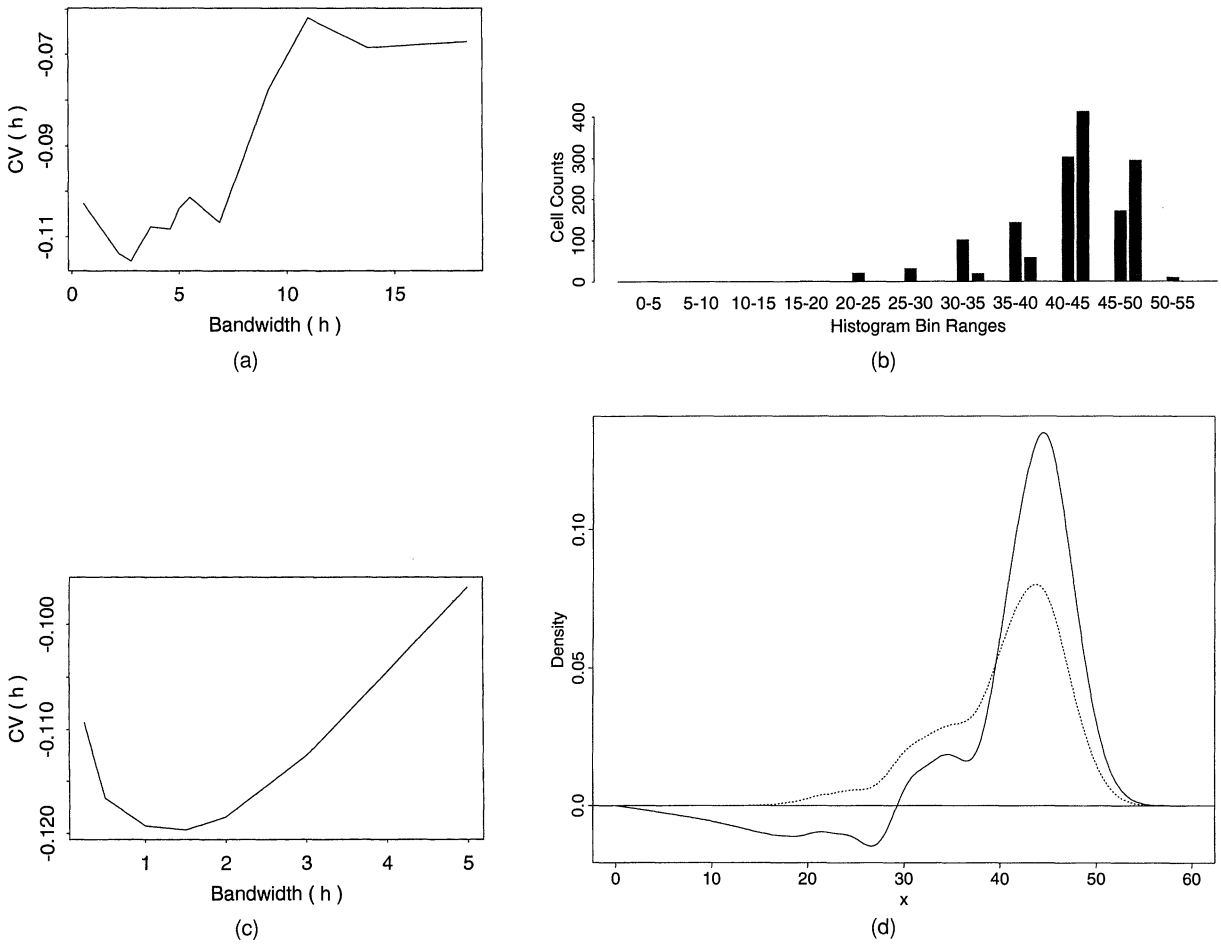

Figure 4. Analysis of a Dataset Comprising 786 Radii (in nm) of Synaptic Terminals From the Electric Organ of the Electric Ray Torpedo marmorata, With Section Thickness of 50 nm. (a) Cross-validation curve for the implicit histogram estimator; (b) barplot of two histograms: for observed (left), and implicitly corrected (right) cell counts; (c) cross-validation curve for the explicit inversion density estimator; (d) density estimate for raw data (dotted curve) and the explicit inversion density estimate (solid curve).

Note that, by (10), the estimator $\hat{f}$ may be written as $C_1\hat{g}$ plus a more complex term, expressible as an integral. It turns out that the variance of the latter term is negligible relative to that of $C_1\hat{g}$, and it is easily shown that the variance of $\hat{g}$ is asymptotic to $(nh)^{-1}\kappa_1 g$. This explains the origins of formula (12). On the other hand, the term of size h^2 in the bias formula (11) has contributions from both parts of the right side of (10); these are $\frac{1}{2}C_1h^2\kappa_2 g''(x)$ and $\frac{1}{2}h^2\kappa_2\{f''(x) - C_1g''(x)\}$.

The variance result in Theorem 1 is very different from its analog in the context of Wicksell's original (thin-slice) problem, where the variance is of size $(nh^2)^{-1}$ rather than, as in (12), $(nh)^{-1}$. This reflects the fact that the thin-slice problem is more ill-posed than its thick-sliced counterpart; see the discussion following (4). The inferior convergence rate in the thin-slice case is a function of the problem, not of the particular estimator type, because it is minimax-optimal (Hall and Smith 1988).

Theorem 1 implies that under conditions (C), the mean integrated squared error is given by $\int E(\hat{f} - f)^2 \sim A_1h^4 + A_2(nh)^{-1}$, where $A_1 = (1/4)\kappa_2^2\int(f'')^2$ and $A_2 = C_1^2\kappa_1$. Therefore, the mean integrated squared error will be asymptotically minimized by taking $h = h_0 \equiv A_3n^{-1/5}$, where $A_3 = \{A_2/(4A_1)\}^{1/5}$. Our next result shows that the cross-validation procedure suggested in Section 2.4 for explicit inversion in the thick-slice case produces a bandwidth that is asymptotically optimal in the sense given here.

For the sake of simplicity, take $w \equiv 1$. Given $0 < \varepsilon < 1/5$, let $\mathcal{H} = \mathcal{H}(n) = [n^{-\varepsilon}, n^{-1+\varepsilon}]$, and let $\hat{h}$ denote the bandwidth in $\mathcal{H}$ that minimizes $\tilde{\gamma}(h)$, defined in Section 2.4.

Theorem 2. Assume those conditions in (C) that pertain to f and K, and in addition suppose that K is Hölder continuous. Then $\hat{h}/h_0 \to 1$ in probability as $n \to \infty$.

Similarly, our cross-validation procedure may be shown to produce asymptotically optimal bandwidths for explicit inversion estimators in the thin-slice case.

4.2 Implicit Inversion

Our main result in this section is tailored to general implicit-inversion density estimators in the thick-slice case, and shows that the approach suggested in Section 2.1 produces consistent density estimators in such circumstances. The main prerequisite is that for each sample size n, the class of potential estimators in $\bar{\mathcal{F}}$ considered when minimizing $\hat{\gamma}(\bar{f})$ be not too rich. In condition (c) of Theorem 3, we describe richness through the number of elements of the class, but alternative approaches could be used. In particular, we could describe it in terms of an increasingly (with n) rich class of "basis" elements, together with a continuum of other elements, each of which is not too far from some element of the basis.

Let $\bar{\mathcal{F}}_n$ denote a subset of $\bar{\mathcal{F}}$, its size assumed to diverge with n. Define

$$\phi_n = \sup_{\bar{f}\in\bar{\mathcal{F}}_n} \sup_{0<x<\infty} |\bar{f}(x) - f(x)|$$

and

$$\psi_n = \inf_{\bar{f}\in\bar{\mathcal{F}}_n} \int (\bar{f} - f)^2,$$

take $w \equiv 1$, and let $\hat{f}$ be the element of $\bar{\mathcal{F}}_n$ that minimizes $\hat{\gamma}(\bar{f})$ over $\bar{f} \in \bar{\mathcal{F}}_n$. We assume the context of the thick-slice Wicksell problem, and assume that if the mean sphere diameter, m, is unknown, then it is estimated root-n consistently.

Theorem 3. Suppose that (a) the supports of f and all elements of $\bar{\mathcal{F}}_n$, for $n \geq 1$, lie within a common compact interval; (b) for some $\delta > 0, \phi_n = O(n^{-\delta})$; and (c) the number N_n of elements of $\bar{\mathcal{F}}_n$ satisfies $N_n = O\{\exp(Cn\psi_n/\log n)\}$ for all $C > 0$. Then

$$\frac{\int(\hat{f} - f)^2}{\inf_{\bar{f}\in\bar{\mathcal{F}}_n}\int(\bar{f} - f)^2} \to 1$$

in probability as $n \to \infty$.

To delineate the implications of this theorem, we outline its specialization to the cases of histograms and frequency polygons. One may show that, provided that f has two derivatives on $(0, \infty), \psi_n = h^2$ or h^4, according to whether $\bar{f}$ is a histogram or its linear interpolant, where $h = h(n)$ denotes binwidth. Take $j = 1, 2$ in these respective cases, suppose that $h \sim c_1n^{-c_2}$, and allow each bin height to assume any one of at most n^{c_3} different values, for constants $c_1, c_2, c_3 > 0$. Then the number of elements of $\bar{\mathcal{F}}_n$ is of order $\exp(c_4h^{-1}\log n)$ for some $c_4 > 0$, and this is no larger than $O\{\exp(Cn\psi_n/\log n)\}$, provided that we insist that $c_2 < 1/(2j + 1)$. (No condition is needed on the value of c_3.) Therefore, the assumptions that $h \sim c_1n^{-c_2}$ with $c_2 < 1/(2j + 1)$, and that the number of possible heights for each bin is no more than polynomially large in n, are sufficient to ensure that the crucial condition (c) holds.

Under conditions similar to those given in Theorem 3, one may prove that our cross-validation method produces consistent implicit-inversion density estimators in the thick-slice case, although for the sake of brevity we do not give details here. Versions of these results in the thin-slice case may also be derived.

APPENDIX: PROOFS

Outline Proof of Theorem 1

We confine attention to the variance formula, because the bias formula may be derived more simply. Define $K_h(z) = h^{-1}K(z/h)$, and write (10) as $\hat{f} = C_1(\hat{g} + \mu^{-1}x\hat{l})$, where $\hat{l}(x) = n^{-1}\sum_i l(x, Y_i)$ and

$$l(x, z) = \int_x^\infty b(x, y)K_h(y - z)\,dy.$$

Because K is bounded and compactly supported and f is bounded, there exist constants $B_1, B_2 > 0$ such that $h|E\{K_h(y_1, Y)K_h(y_2 - Y)\}| \leq B_1I(|y_1 - y_2| \leq B_2h)$. Hence

$$\begin{aligned} E\{l(x, Y)^2\} &\leq h^{-1}B_1\int_x^\infty\int_x^\infty b(x, y_1)b(x, y_2) \\ &\quad \times I(|y_1 - y_2| \leq B_2h)\,dy_1\,dy_2 \\ &= o(h^{-1}) \end{aligned}$$

as $h \to 0$. Hence $\text{var}\ \hat{l} = o_p\{(nh)^{-1}\}$, and so $\text{var}\ \hat{f} \sim C_1^2\ \text{var}\ \hat{g} \sim (nh)^{-1}C_1^2\kappa_1 g$. The central limit theorem for $\hat{f}$ may be proved via Lindeberg's theorem.

Outline Proof of Theorem 2

We assume for simplicity that m is known rather than estimated from the data. The latter case may be treated by incorporating a relatively simple, subsidiary argument. We prove that

$$n^{-1}\sum_{i=1}^{n} L(\hat{f}_{hi})(Y_i) = \int E(\hat{f}_h)f + R_1 + o_p\{(nh)^{-1} + h^4\} \quad \text{(A.1)}$$

uniformly in $h \in \mathcal{H}$, where R_j denotes a random variable not depending on h. Similarly, it may be shown that $\int \hat{f}_h^2 = \int E(\hat{f}_h^2) + o_p\{(nh)^{-1} + h^4\}$ uniformly in $h \in \mathcal{H}$. Together, these results imply that

$$\begin{aligned}\tilde{\gamma}(h) &\equiv \int \hat{f}_h^2 - 2n^{-1}\sum_{i=1}^{n} L(\hat{f}_{hi})(Y_i)\\ &= \int E(\hat{f}_h - f)^2 + R_2 + o_p\{(nh)^{-1} + h^4\}\end{aligned}$$

uniformly in $h \in \mathcal{H}$. Theorem 2 then follows from this formula and from the expansion of $\int E(\hat{f}_h - f)^2$ implied by Theorem 1.

Put $K_h(x) = h^{-1}K(x/h)$ and let $A(y|h)$ denote the functional $\{A(y|h)\}(x) = K_h(x-y)$. Then $\hat{f}_{hi} = (n-1)^{-1}\sum_{j:j\neq i} A(Y_j|h)$, and so

$$L(\hat{f}_{hi}) = (n-1)^{-1}\sum_{j:j\neq i} L\{A(Y_j|h)\},$$

whence

$$\begin{aligned}\hat{J}(h) &\equiv \frac{1}{2}n^{-1}\sum_{i=1}^{n} L(\hat{f}_{hi})(Y_i)\\ &= \{n(n-1)\}^{-1}\sum_{1\le j<i\le n}\sum L\{A(Y_j|h)\}(Y_i).\end{aligned}$$

Using the definition of β_α at (6), define

$$J = \{n(n-1)\}^{-1}\sum_{i=1}^{n}(i-1)C\beta_g(Y_i)$$

and

$$\begin{aligned}Z_{i1} &= \sum_{j=1}^{i-1}\beta_{A(Y_j|h)}(Y_i) - (i-1)\beta_g(Y_i)\\ &= (i-1)\left\{u(0)D_i(y) - \int_0^{Y_i} a_{10}(x, Y_i)D_i(x)\,dx\right\}, \quad \text{(A.2)}\end{aligned}$$

where $D_i(y) = (i-1)^{-1}\sum_{j\le i-1} A(Y_j|h)(y) - g(y)$. Then

$$C^{-1}\{\hat{J}(h) - J\} = \{n(n-1)\}^{-1}\sum_{i=2}^{n} Z_{i1}. \quad \text{(A.3)}$$

Define $\lambda_i = \{(i-1)h\}^{-1/2} + h^2$ for $2 \le i \le n$, and observe that by Rosenthal's inequality (Hall and Heyde 1980, p. 23),

$$\sup_{2\le i\le n<\infty}\ \sup_{h\in\mathcal{H}} \lambda_i^{-2k}\ \sup_{0<y<\infty} E\{D_i(y)^{2k}\} < \infty$$

for all integers $k \ge 1$. Therefore, by (A.2), $E\{(i-1)^{-1}Z_{i1}\}^{2k} = O(\lambda_i^{2k})$ uniformly in $2 \le i \le n < \infty$ and $h \in \mathcal{H}$. Hence

$$\begin{aligned}E\left[\left\{\sum_{i=2}^{n} E(Z_{i1}^2|Y_1,\ldots,Y_{i-1})\right\}^k\right] &\le \left[\sum_{i=2}^{n}\{E(Z_{i1}^{2k})\}^{1/k}\right]^k\\ &\le B_1(k)\left(\sum_{i=2}^{n} i^2\lambda_i^2\right)^k\\ &\le B_2(k)[n^3\{(nh)^{-1} + h^4\}]^k \quad \text{(A.4)}\end{aligned}$$

and

$$\sum_{i=2}^{n} E(Z_{i1}^{2k}) \le B_3(k)n^{2k+1}\{(nh)^{-1} + h^4\}^k, \quad \text{(A.5)}$$

where $B_1(k), B_2(k), \ldots$ denote positive constants depending only on f, μ, and k. Put $Z_{i2} = E(Z_{i1}|Y_1,\ldots,Y_{i-1})$ and $Z_{i3} = Z_{i1} - Z_{i2}$. Then $E(Z_{i3}|Y_1,\ldots,Y_{i-1}) = 0$, and so the Z_{i3}'s are martingale differences. Therefore, by Rosenthal's inequality together with (A.4) and (A.5),

$$E\left(\sum_{i=2}^{n} Z_{i3}\right)^{2k} \le B_4(k)n^{3k}\{(nh)^{-1} + h^4\}^k. \quad \text{(A.6)}$$

Let Y denote a generic Y_j, and define

$$\begin{aligned}Z_1(Y) &= u(0)\int g(Y - hz)K(z)\,dz\\ &\quad - \int_0^\infty g(y)\,dy\int_{\{z:Y-hz\le y\}} a_{10}(Y - hz, y)K(z)\,dz\\ &= -\int (Y - hz)K(z)\,dz\\ &\quad \times \int_{Y-hz}^\infty a(Y - hz, y)\{y^{-1}g(y)\}'\,dy \quad \text{(A.7)}\end{aligned}$$

and

$$Z_2(Y) = -Y\int_Y^\infty a(Y, y)\{y^{-1}g(y)\}'\,dy, \quad \text{(A.8)}$$

where the prime denotes differentiation. Put $Z_3 = Z_1(Y) - Z_2(Y)$ and $Z_{i,l+2} = \sum_{j\le i-1} Z_l(Y_j)$ for $l = 2, 3$. Given a random variable W, write "$W-$ mean" for "$W - E(W)$." Note that Z_2 does not depend on h, and so

$$\begin{aligned}\sum_{i=2}^{n} Z_{i2} - \text{ mean} &= \sum_{i=2}^{n}\sum_{j=1}^{i-1} Z_1(Y_j) - \text{ mean}\\ &= \sum_{j=1}^{n-1}(n-j)Z_3(Y_j) - \text{ mean} + R_3. \quad \text{(A.9)}\end{aligned}$$

Moreover, by Rosenthal's inequality,

$$\begin{aligned}&E\left\{\sum_{j=1}^{n-1}(n-j)Z_3(Y_j) - \text{ mean}\right\}^{2k}\\ &\quad \le B_5(k)[n^{3k}\{EZ_3(Y)^2\}^k + n^{2k+1}E\{Z_3(Y)^{2k}\}]. \quad \text{(A.10)}\end{aligned}$$

We claim that

$$E\{Z_3(Y)^{2k}\} \le B_6(k)h^{4k}. \quad \text{(A.11)}$$

Combining (A.9)–(A.11), we deduce that

$$E\left(\sum_{i=2}^{n} Z_{i2} - \text{ mean} - R_3\right)^{2k} \le B_7(k)n^{3k}h^{4k},$$

whence by (A.3) and (A.6), noting that J does not depend on h,

$$E\{\hat{J}(h) - \text{ mean } - R_4\}^{2k} \le B_8(k)n^{-k}\{(nh)^{-1} + h^4\}^k. \quad \text{(A.12)}$$

In view of (A.12), there exists $\delta > 0$ (depending on ε in the definition of $\mathcal{H}$) such that

$$\sup_{h\in\mathcal{H}} E\left(\frac{\hat{J}(h) - E\hat{J}(h) - R_4}{n^{-\delta}\{(nh)^{-1} + h^4\}}\right) = O(n^{-\delta k})$$

for all k. Therefore, if $\mathcal{H}'$ denotes a subset of $\mathcal{H}$ containing $O(n^B)$ elements for some $B > 0$, then, by Markov's inequality,

$$P[|\hat{J}(h) - E\hat{J}(h) - R_4| > n^{-\delta}\{(nh)^{-1} + h^4\} \quad \forall\, h \in \mathcal{H}'] \to 0.$$

Because B is arbitrary, the Hölder continuity assumption enables us to replace $\mathcal{H}'$ by $\mathcal{H}$ in this result. Formula (A.1), with R_1 there equal to $2R_4$, follows immediately.

It remains to prove (A.11). Let γ be a function with a bounded derivative. Because $a_{10}(x,y) = -(x/y)a_{01}(x,y)$, then, through integration by parts,

$$\int_x^\infty a_{10}(x,y)\gamma(y)\,dy = \gamma(x)u(0) + x\int_x^\infty a(x,y)\{y^{-1}\gamma(y)\}'\,dy.$$

Hence, using the exact formula for the remainder in a Taylor expansion,

$$\begin{aligned}
&\int_{x+\eta}^\infty a(x+\eta,y)\gamma(y)\,dy - \int_x^\infty a(x,y)\gamma(y)\,dy \\
&\quad = \eta\int_0^1 (x+\eta t)\,dt\int_{x+\eta t}^\infty a(x+\eta t,y)\{y^{-1}\gamma(y)\}'\,dy \\
&\quad = \eta\int_0^1 (x+\eta t)\,dt\int_x^\infty a(x,y)\{y^{-1}\gamma(y)\}'\,dy + O(\eta^2)
\end{aligned} \quad \text{(A.13)}$$

as $\eta \to 0$, uniformly in $x \in [x_1, x_2]$ for any $0 < x_1 < x_2 < \infty$. Take $x = Y, \eta = -hz$, and $\gamma(y) = \{y^{-1}\gamma(y)\}'$; multiply both sides of (A.13) by $(Y - hz)K(z)$; and integrate over z, obtaining

$$\text{ess. sup}|Z_1(Y) - Z_2(Y)| \le \text{ const. } h^2.$$

[Note formulas (A.7) and (A.8).] This implies (A.11).

Outline Proof of Theorem 3

We assume initially that m is known, and at the end of our proof address the case where it must be estimated. Put

$$S = S(\bar{f}) = n^{-1}\sum_{i=1}^n \{L(\bar{f}) - L(f)\}(Y_i)$$

and define $\phi = \phi(\bar{f}) = \sup_x |\bar{f}(x) - f(x)|$ and $\psi = \psi(\bar{f}) = \int(\bar{f} - f)^2$. Note from (6) that

$$\begin{aligned}
D(y) \equiv C^{-1}\{L(\bar{f}) - L(f)\}(y) \\
= (\bar{f} - f)(y)u(0) - \int_0^y a_{10}(x,y)(\bar{f} - f)(x)\,dx.
\end{aligned}$$

Furthermore, $|a_{10}(x,y)| \le C_1 x(y^2 - x^2)^{-1/2}$, where $C_1, C_2, \ldots$ denote positive constants not depending on $x, y, \bar{f}$ or n. Hence, because f and $\bar{f}$ are both supported within a given compact interval, $|D(y)| \le C_2\phi$ and, for each $3/2 < p < 2$ and $q = (1 - p^{-1})^{-1}$,

$$\text{var}\{D(Y)\} \le C_3\left\{\psi + (2-p)^{-2/p}\left(\int |\bar{f} - f|^q\right)^{2/q}\right\}.$$

(We used Hölder's inequality to derive the last term.) Therefore, taking $p = 2 - (\log n)^{-1}$, we have by assumption (b) that $\text{var}\{D(Y)\} \le C_4\psi\log n$. From these results and Bernstein's inequality (e.g., Pollard 1984, p. 193), we may prove that for each $t > 0$,

$$\begin{aligned}
\pi_1(t) \equiv P\{|S - E(S)| > t\} \\
\le 2\exp\{-C_5 nt^2/(\psi\log n + \phi t)\}.
\end{aligned} \quad \text{(A.14)}$$

Take $t = \psi\eta$, where $\eta = \eta(n)$ denotes a sequence of positive constants converging to 0. Then $\psi\log n \ge \phi t$ for all sufficiently large n, and so by (A.14), $\pi_1(t) \le 2\exp(-C_6 n\psi\eta^2/\log n)$. Hence

$$\begin{aligned}
\pi_2(n) &\equiv P\{\sup_{\bar{f}\in\bar{\mathcal{F}}_n} \psi(\bar{f})^{-1}|S(\bar{f}) - ES(\bar{f})| > \eta\} \\
&= O\{N_n\exp(-C_6 n\psi_n\eta^2/\log n)\}.
\end{aligned} \quad \text{(A.15)}$$

By assumption, $n\psi_n/\log n \to \infty$. Because $N_n = O\{\exp(Cn\psi_n/\log n)\}$ for all $C > 0$, then, provided that $\zeta = \zeta(n) \to 0$ sufficiently slowly, we have

$$n\zeta^2\psi_n/\log n \to \infty$$

and

$$N_n = O\{\exp(n\zeta^2\psi_n/\log n)\}.$$

For this ζ, taking $\eta = \zeta^{1/2}$ in (A.15), we obtain

$$\pi_2(n) = O[\exp\{(n\zeta^2\psi_n/\log n) - (C_6 n\zeta\psi_n/\log n)\}] \to 0.$$

Hence

$$S(\bar{f}) = ES(\bar{f}) + o_p\{\psi(\bar{f})\} \quad \text{(A.16)}$$

uniformly in $\bar{f} \in \bar{\mathcal{F}}_n$. Now $ES(\bar{f}) = I(\bar{f}, f) - I(f, f)$, and so, uniformly in $\bar{f} \in \bar{\mathcal{F}}_n$,

$$\hat{\gamma}(\bar{f}) = \int \bar{f}^2 - 2S(\bar{f}) - 2\hat{I}(f) = \psi(\bar{f}) - R(f) + o_p\{\psi(\bar{f})\},$$

where $R(f) = 2I(f,f) + 2\hat{I}(f) + \int f^2$ does not depend on $\bar{f}$. In the case where m is known, the theorem follows from this formula.

If m is not known, then it is replaced by a root-n–consistent estimator, $\hat{m}$, in the definitions of $C = C(m)$ and $\hat{I}(\bar{f}) = C(m)n^{-1}\sum_i L(\bar{f})(Y_i)$. Arguing in this way, we may show, using (A.16), that

$$\begin{aligned}
\hat{I}(\bar{f}) &= C(\hat{m})C(m)^{-1}S(\bar{f}) + C(\hat{m})n^{-1}\sum_{i=1}^n L(f)(Y_i) \\
&= C(\hat{m})C(m)^{-1}I(\bar{f}, f) + R_1(\bar{f}),
\end{aligned} \quad \text{(A.17)}$$

where $R_j(\bar{f})$ denotes terms that either do not depend on $\bar{f}$ or are $o_p\{\psi(\bar{f})\}$ uniformly in $\bar{f} \in \bar{\mathcal{F}}_n$. In view of the root-$n$ consistency of $\hat{m}$, $C(\hat{m})C(m)^{-1} = 1 + O_p(n^{-1/2})$. Moreover, $I(\bar{f}, f) - I(f, f) = O\{\psi(\bar{f})^{1/2}\}$. Because $n^{-1/2}\psi(\bar{f})^{1/2} = o\{\psi(\bar{f})\}$ uniformly in $\bar{f} \in \bar{\mathcal{F}}_n$, then, by (A.17), $\hat{I}(\bar{f}) = I(\bar{f}, f) + R_2(\bar{f})$, whence it follows that $\hat{\gamma}(\bar{f}) = \psi(\bar{f}) + R_3(\bar{f})$ uniformly in $\bar{f} \in \bar{\mathcal{F}}_n$. The theorem is a consequence of this result.

[Received June 1998. Revised April 1999.]

REFERENCES

Anderssen, R. S. (1980), "On the Use of Linear Functionals for Abel-Type Integral Equations in Applications," in *Application and Numerical Solution of Integral Equations*, eds. R. S. Anderssen, F. R. de Hoog, and M. A. Lukas, The Hague: Nijhoff, pp. 195–221.

Anderssen, R. S., and de Hoog, F. R. (1990), "Abel Integral Equations," in *Numerical Solution of Integral Equations*, ed. M. A. Goldberg, New York: Plenum, pp. 373–410.

Anderssen, R. S., and Jakeman, A. J. (1974), "On Computational Stereology," in *Proceedings of the Sixth Australian Computing Conference, Sydney*, Sydney: Australian Computer Society, pp. 353–362.

——— (1975a), "Abel-Type Integral Equations In Stereology, I: General Discussion," *Journal of Microscopy*, 105, 121–133.

——— (1975b), "Abel-Type Integral Equations in Stereology, II: Computational Methods of Solution and the Random Spheres Approximation," *Journal of Microscopy*, 105, 135–153.

Bach, G. (1967), "Kugelgrößenverteilung und Verteilung der Schnittkreise, Ihre Wechselseitgen Beziehungen und Verfahren zur Bestimmung der Einen aus der Anderen," in *Proceedings of the Symposium on Quantitative Methods in Morphology*, eds. E. R. Weibel and H. Elias, Berlin: Springer, pp. 23–45.

——— (1976), "Über die Auswertung von Schnittflächenverteilungen," *Biometrical Journal*, 18, 407–412.

Baddeley, A. (1982), "Stochastic Geometry: An Introduction and Reading List," *Internat. Statist. Rev.*, 50, 179–193.

Becker, R. A., Chambers, J. M., and Wilks, A. R. (1988), *The New S Language*, Pacific Grove, CA: Wadsworth & Brooks/Cole.

Bowman, A. W. (1984), "An Alternative Method of Cross-Validation for the Smoothing of Density Estimates," *Biometrika*, 71, 353–360.

Carroll, R. J., van Rooij, A. C. M., and Ruymgaart, F. H. (1991), "Theoretical Aspects of Ill-Posed Problems in Statistics," *Acta Appl. Math.*, 24, 113–140.

Fox, G. Q. (1988), "A Morphometric Analysis of Synaptic Vesicle Distributions," *Brain Research*, 475, 103–117.

Goldsmith, P. L. (1967), "The Calculation of True Particle Size Distributions From the Sizes Observed in a Thin Slice," *British J. of Applied Physics*, 18, 813–830.

Grenander, U. (1981), *Abstract Inference*, New York: Wiley.

Hall, P. (1988), *Introduction to the Theory of Coverage Processes*, New York: Wiley.

Hall, P., and Heyde, C. C. (1980), *Martingale Limit Theory and Its Application*, New York: Academic Press.

Hall, P., and Smith, R. L. (1988), "The Kernel Method for Unfolding Sphere Size Distributions," *J. Computat. Phys.*, 74, 409–421.

Hoogendoorn, A. W. (1992), "Estimating the Weight Undersize Distribution for the Wicksell Problem," *Statistica Neerlandica*, 4, 259–282.

Jakeman, A. J. (1984), "Numerical Inversion of a Second Kind Singular Volterra Equation—The Thin Section Equation of Stereology," *Utilitas Mathematica*, 26, 193–213.

Jakeman, A. J., and Anderssen, R. S. (1974), A Note on Numerical Methods for the Thin Section Model," *Electron Microscopy*, 2, 4–5.

Johnstone, I. M., and Silverman, B. W. (1991), "Discretization Effects in Statistical Inverse Problems," *Journal of Complexity*, 7, 1–34.

Mase, S. (1995), "Stereological Estimation of Particle Size Distributions," *Adv. Appl. Probab.*, 27, 350–366.

Mecke, J., and Stoyan, D. (1980), "Stereological Problems for Spherical Particles," *Mathematische Nachrichten*, 96, 311–317.

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

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

Rudemo, M. (1982), "Empirical Choice of Histograms and Kernel Density Estimators," *Scandinavian Journal of Statistics*, 9, 65–78.

Ruymgaart, F. H. (1993), "A Unified Approach to Inversion Problems in Statistics," *Mathematical Methods of Statistics*, 2, 130–146.

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

Taylor, C. C. (1983), "A New Method for Unfolding Sphere Size Distributions," *Journal of Microscopy*, 132, 57–66.

van Es, B., and Hoogendoorn, A. (1990), "Kernel Estimation in Wicksell's Corpuscle Problem," *Biometrika*, 77, 139–145.

Weibel, E. R. (1980), *Stereological Methods, Vol. 2: Theoretical Foundations*, London: Academic Press.

Wicksell, S. D. (1925), "The Corpuscle Problem, Part I," *Biometrika*, 17, 84–99.